BREAST CANCER SUBTYPES, MOUSE MODELS, AND MICROARRAYS

Jason I. Herschkowitz

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Genetics and Molecular Biology

Chapel Hill
2007

Approved by:

Charles Perou
Norman Sharpless
David Threadgill
Channing Der
Shelton Earp

ABSTRACT

JASON I. HERSCHKOWITZ: Breast Cancer Subtypes, Mouse Models, and Microarrays
(Under the direction of Charles Perou)

Breast cancer can no longer be viewed as a single disease. Molecular profiling studies have altered the way we consider breast cancer, showing us that there are several subtypes, each with their own unique biology. There are many model systems available in which to study breast cancer, however, each of these comes with advantages and disadvantages. We chose the mouse as a model to investigate breast cancer biology because it gives us the ability to study tumor progression and response to therapy *in vivo*. Numerous mouse models of breast carcinomas have been developed. The extent to which any faithfully represent clinically significant human phenotypes was unknown. Analogous to our human studies, we characterized mammary tumor gene expression profiles from a large number of murine models using DNA microarrays and compared the resulting data to our human breast tumor dataset. Two major applications of across-species tumor comparisons surfaced from these studies. First, we were able to determine that mouse models contain many of the global characteristics of particular classes or subtypes of human tumors. This included basal versus luminal distinctions, a proliferation/cell cycle signature, and a fibroblast signature. Second, the mouse models were able to inform the human disease; for example, we identified an amplicon that included the K-ras gene present in both mouse and human basal tumors. The high proliferation seen in common between mouse models of Rb loss and human basal-like breast tumors hinted that there is an Rb defect in this human subtype. And finally the mouse

spindloid tumors shared significant gene overlap with a new molecular subtype of breast cancer. Although no single murine model recapitulated all the expression features of a given human subtype, these shared expression features have provided us a common framework so that we can now integrate these murine mammary tumor models into our studies of human breast cancer.

ACKNOWLEDGEMENTS

I would like to thank:

My advisor, Chuck Perou

The past and present members of the Perou Lab:

Katie Hoadley, Melissa Troester, Xiaping He, Olga Karginova, Aaron Thorner,

Victor Weigman, George Chao, George Murrow, Joel Parker, Cheng Fan, Jerry Usary,

David Darr, Zhiyuan Hu, Mei Liu, Okey Ukairo

My committee members:

Norman Sharpless
David Threadgill
Channing Der
Shelton Earp

My wife: Rebecca Herschkowitz

My children: Avi, Julien, and Lilia Herschkowitz

My parents: Renee and Stuart Herschkowitz

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

ABBREVIATIONS

DCIS	ductal carcinoma in situ
DMBA	7,12-dimethylbenzanthracene
EMT	epithelial to mesenchymal transition
ER	estrogen receptor
GEM	genetically engineered mouse
GO	gene ontology
GSEA	gene set enrichment analysis
IBC	inflammatory breast cancer
IDC	invasive ductal carcinoma
IHC	immunohistochemistry
ILC	invasive lobular carcinoma
LOH	loss of heterozygosity
MMTV	mouse mammary tumor virus
PyMT	polyoma middle T-antigen
RB	retinoblastoma
SAM	significance analysis of microarrays
TAG	T-antigen
VNTR	variable number tandem repeat
WAP	whey acid protein

CHAPTER I

INTRODUCTION

In women in the United States, breast cancer is currently the most commonly diagnosed tumor type, and the second leading cause of cancer death[1]. One in eight American women will develop this disease in their lifetime. Some of the non-modifiable risk factors include age, family history, age at first full term pregnancy, early menarche, and late menopause. Modifiable risk factors include postmenopausal obesity, use of postmenopausal hormones, alcohol consumption, and physical inactivity. From cell lines to xenografts, there are many model systems to study breast cancer[2]. Each of these model systems comes with inherent advantages and disadvantages. Different models may represent distinct aspects or subtypes of this heterogeneous disease and studying one model will never be adequate to encompass the whole disease. My research has focused on the mouse because we desired a model that will give us the ability to study tumor formation, progression and therapy response *in vivo*. Furthermore, the mouse is amenable to genetic manipulation. There are many mouse mammary tumor models that have been generated over the past twenty years. In order to use these models in the most relevant way, the question we first need to answer is how do these mouse models relate to the human disease and more specifically, what tumor subtypes of breast cancer are represented by each model? It is these models that most closely mimic the human disease that will be the most useful in further studies.

Breast cancer subtypes

Classification of breast cancer into molecular subtypes using gene expression profiling has drastically changed our understanding of breast cancer biology[3-5]. Breast cancer can now be divided into at least five subtypes. The biology of these subtypes is indicative of different cell types of origin. Two of these subtypes are estrogen receptor (ER) negative: High expression of basal cytokeratins 5 and 17, reminiscent of normal breast basal/myoepithelial cells, is characteristic of the basal-like subtype. HER2+/ER- tumors, which are the second ER- subtype, generally overexpress HER2 and often show amplification of this locus. Two of these subtypes are ER-positive, with the Luminal A tumors expressing an important quartet of transcription factors (ER, GATA3, XBP1, and FOXA1)[6] as well as luminal cytokeratins 8 and 18. Luminal B tumors express the same luminal cell type genes at a lower level and also show higher expression of proliferation genes than luminal A tumors. The normal-like subtype contains breast tumors that cluster together with normal breast samples; this is thought to be due to mis-sampling of tumor tissue caused by contamination with too much normal mammary gland tissue.

Molecular subtypes and outcome

Molecular subtypes are predictive of clinical outcome with the Luminal A subtype having a generally more favorable outcome[3]. Individuals with HER2+/ER- and basal-like tumors generally have the worst outcomes. Studies have also shown that these different molecular subtypes have distinct responses to therapy[7, 8]. The basal-like subtype has drawn a lot of attention because these patients generally have a poor outcome and unlike the other subtypes that have targeted therapies with measurable efficacies in current clinical use

(luminal A and B – tamoxifen and aromatase inhibitors, HER2+ - trastuzamab), the basal-like subtype does not have any efficacious targeted therapy. In addition, the breast cancer molecular subtypes have different patterns of metastasis. In a recent study it was shown that brain-metastasizing breast cancer belongs predominantly to the basal-like and HER2+ subtypes[9], while Luminal A tumors tend to metastasize to the bone.

Several recent studies have examined ductal carcinoma in situ (DCIS) lesions and determined that they display immunophenotypes analogous to the subtypes of invasive breast carcinomas [10-13]. Interestingly, basal-like DCIS was associated with unfavorable prognostic variables including high-grade nuclei, overexpression of p53 (an indication of p53 mutation), and a high Ki-67 labeling index[12]. These studies show that potential precursor lesions can be identified that give rise to the breast cancer subtypes The classifications of DCIS have important implications given the increased numbers of DCIS currently diagnosed with the use of mammography.

Genetic changes and subtype

Each breast cancer subtype has specific associated genetic changes (Table 1.1). p53 mutations have been shown to be common in basal-like, HER2+/ER-, and luminal B tumors, but infrequent in the luminal A subtype[3]. BRCA1 carriers have been shown to be predisposed to developing basal-like tumors[4]. Phosphatidylinositol 3-kinase, catalytic, α polypeptide (PIK3CA) mutations are frequent in the luminal A subtype (unpublished and [14]). Recurrent copy number alterations have also been shown to differ between tumor subtypes[15, 16]. Basal-like tumors are relatively enriched for low-level copy number gains

and losses. High-level DNA amplification on the other hand has been shown to be very frequent in luminal B tumors including the region on chromosome 8q containing the c-myc oncogene as well as 11q containing Cyclin D1. As expected, HER2+ER- tumors have been shown to have high-level amplification of 17q harboring HER2/ERBB2. Subtypes of breast cancer also show signature loss of heterozygosity (LOH) events[17]. X-chromosomal isodisomy, the loss of the inactive X and duplication of the active X, has been shown to be characteristic of basal-like tumors[18].

Table 1.1	Alterations and risk factors associated with the molecular subtypes of breast cancer				
Alteration	**Luminal A**	**Luminal B**	**HER2+/ER-**	**Basal-like**	**References**
Familial				BRCA1	[4]
Mutations					
p53	few	+	++	+++	[3]
PIK3CA	+++	++	+	++	unpublished
Hormone receptors	ER+, PR+, AR+	ER+, PR+, AR+	ER-, PR-, AR+	ER-, PR-, AR-	
Copy Number					
Gains	**1q, 16p**	1p, 1q, 8q, 17q, 20p, 20q	1q, 7p, 8q, 16p, 20q	1q, 3q, 6q, 7q, 8q, **10p,** 11p, 17q, 21q	[15, 16]
Losses	16q	1p, 3p, 3q, 8p, 13q, 16q, 17p, 22q	1p, 8p, 13q, 18q	**3p, 4p, 4q, 5q, 12q,** 13q**, 14q, 15q**	[18]
Amplifications	8p11-12, 11q13-14, 12q13-14, 17q11-12, 17q21-24, 20q13	6q, 7p, 8p11-12, two regions **8q**, 9q, 11q13-14, 19q, 20q	**17q**		
Other				X-chromosomal isodisomy	
Risk factors and associations	old age, pregnancy protective	old age, alcohol, HRT use		early age, early age at menarche, AA, high WHR, high BMI, pregnancy risk, pregnancy-no lactation risk	[19-21]

Bold indicates changes found in both copy number studies.

Breast cancer subtypes and risk factors

In a recent study, the prevalence of breast cancer subtypes was shown to vary by race[19]. The basal-like breast cancer subtype was more prevalent among premenopausal African American women compared with postmenopausal African American women and non–African American women of any age in the Carolina breast cancer study. The luminal A subtype, on the other hand, was less prevalent among premenopausal African American women. Recently, it has been discovered that risk factors differ between the subtypes in large population based studies[20, 21]. In the Carolina Breast Cancer Study (CBCS), pregnancy was shown to be protective for luminal A tumors with increased parity, and younger age at first full-term pregnancy each showing a reduced risk. A high waist-to-hip ratio increased the risk of luminal A tumors in postmenopausal women. On the other hand, pregnancy was shown to be associated with basal-like breast cancer with parity and younger age at first full-term pregnancy increasing risk. Women who did not breastfeed were also shown to have an increased risk of basal-like breast cancer showing that there are protective affects from lactation. Abdominal adiposity was also associated with basal-like breast cancer in pre-and postmenopausal women in the CBCS study[20]. In the Polish Breast Cancer Study, increasing body mass index reduced the risk of luminal A tumors among premenopausal women[21]. Increasing age at menarche was associated with a reduced risk of basal-like breast cancer in this study. Family history was a risk factor in all the subtypes, but had the greatest impact in the basal-like subtype. These studies show that the subtypes display distinct risk factor profiles and not surprisingly, luminal A, the most common subtype, shows associations typically described for breast cancer as a whole.

Breast cancer has long been studied and treated as one disease. The molecular profiling and epidemiological studies have shown that breast cancer should be considered as several distinct diseases that develop along a continuum of epithelial cell types in the mammary gland. Molecular genetic and epidemiological studies of these subtypes show that each of these has different etiologies and patterns of progression. This is important when considering preventative and therapeutic strategies. Therefore, no one model should be expected to adequately represent breast cancer and studying multiple models representative of each molecular subtype would be of great benefit.

The mouse as a model of human breast cancer

Beginning with the first transgenic mouse mammary tumor model, the MMTV-Myc mouse[22], developed by Phil Leder and coworkers, researchers have spent the last three decades genetically engineering mouse models to study human breast cancer. There are several advantages to using the mouse as a model including the ability to manipulate their genes and genomes, their short generation times, and small size. They physiologically and molecularly resemble humans more than many other model organisms. Like the human, the mouse genome is also fully sequenced. Mouse models also provide an *in vivo* model system for the evaluation of different therapies. Although xenograft models, human cancer cell lines grown in immunodeficient mice, have had limited success in predicting drug efficacy, early results with genetically engineered mice (GEM) have been promising [23]. Murine models also show some caveats including different mammary physiology and estrous cycle. There are also indications that there is a difference in the capacity of primary cells to be transformed between the two species[24].

In 1999 a panel of distinguished medical and veterinary pathologists with mammary gland expertise gathered in Annapolis, Maryland to develop a classification of breast lesions based on their examination of 39 GEM models[25]. One of the main conclusions of their consensus report was that overall GEMs give rise to tumors that histologically do not resemble the common types of human breast cancer. To date, many studies have been performed where rodent models and human tumors are compared pathologically and phenotypically [26], however, it is still not clear how accurately most murine models mimic human tumors. With the advent of genomic profiling, a new tool for comparative genomics emerged offering a powerful means for cross species comparisons. We hypothesized that some murine mammary models will better mimic specific human tumor subtypes than others. These similarities can be identified using genomic profiling.

Mouse tumor models are created with genetic lesions seen in humans

Murine mammary tumor models have shown us that genes associated with human breast cancer can induce cancer in mice. Inactivation of p53, BRCA1, and Rb, and activation/overexpression of MYC and HER2 have been used to initiate tumorigenesis in the mouse mammary gland. Models have been created through a multitude of approaches including the transgenic overexpression of genes involved in human cancer, the production of dominant negative proteins to disrupt protein function, the targeted disruption of tumor suppressor genes, and by mutagenesis through infection with the Mouse Mammary Tumor Virus (MMTV) or treatment with chemical agents (i.e. DMBA). A number of different rodent gene promoters have been employed in order to achieve mammary specific

expression. A majority of the transgenics generated have employed either the MMTV Long Terminal Repeat (MMTV-LTR) or the whey acidic protein (WAP) promoter. The MMTV-LTR is active in ductal and alveolar cells throughout mammary development and its transcriptional activity increases during pregnancy. In contrast, the WAP promoter, which expresses preferentially in alveolar cells, is only active in the mid-pregnant and lactating mammary gland. The C3(1) component of the rat prostate steroid binding protein has also been used to drive transgene expression in one of the mouse mammary cancer models. A variety of other promoters have been used to drive expression in the mammary gland including H19, bovine β-lactoglobin, metallothionen, keratin 5, and keratin 14. The choice of transgene and promoter are both critical to the successful creation of relevant animal models.

We have chosen to use several models, most of which are commonly used and many of which are initiated by events seen in human breast cancer (Figure 1.1). The models/genes that we used in our initial studies are c-MYC, HER2/Neu, p53, RB, BRCA1, WNT1, and INT3/Notch4. The first gene of interest is the c-myc oncogene, which is amplified in 30% and overexpressed in almost 70% of human breast cancer. Models targeting the c-myc oncogene to the mammary gland of mice have been created using both the MMTV promoter[22] and the WAP promoter[27]. These models have been important in showing that the overexpression of the c-myc oncogene in the mammary gland can initiate tumorigenesis.

The second gene is HER2/ERBB2 (human name), that encodes a receptor tyrosine kinase belonging to the epidermal growth factor receptor (EGFR) family and that has been shown to be amplified in 15-30% of human breast cancers[28-31]. As described above, overexpression of this gene is associated with a specific human subtype (HER2+ subtype) that is ER-negative. It should also be pointed out that a subset of human tumors of the ER-positive luminal subtypes also highly express HER2, suggesting that HER2 tumorigenesis can occur in at least 2 distinct cell types. MMTV has been used to drive expression of HER2, also called ERBB2 or Neu in the mouse mammary gland[32-34](see [35]for an extensive review). These mice develop mammary adenocarcinomas that have been described as histologically and cytologically very similar to human ductal carcinoma-*in situ* (DCIS)[26]. Also, expression of the viral protein polyoma middle T-antigen (PyMT) using the MMTV promoter leads to mammary tumors that are similar to Neu driven cancers[36]. Tumors from this model are said to resemble the common scirrhous carcinomas in humans[26], while papillary tumors produced from this model resemble the less common human papillary breast cancer[26]. Like Neu[37], the PyMT model has also been shown to activate the src and PI3 kinase pathways[38]. Interestingly, these mice also have a high incidence of metastasis to the lung, which is greater than that seen in the MMTV-Neu model.

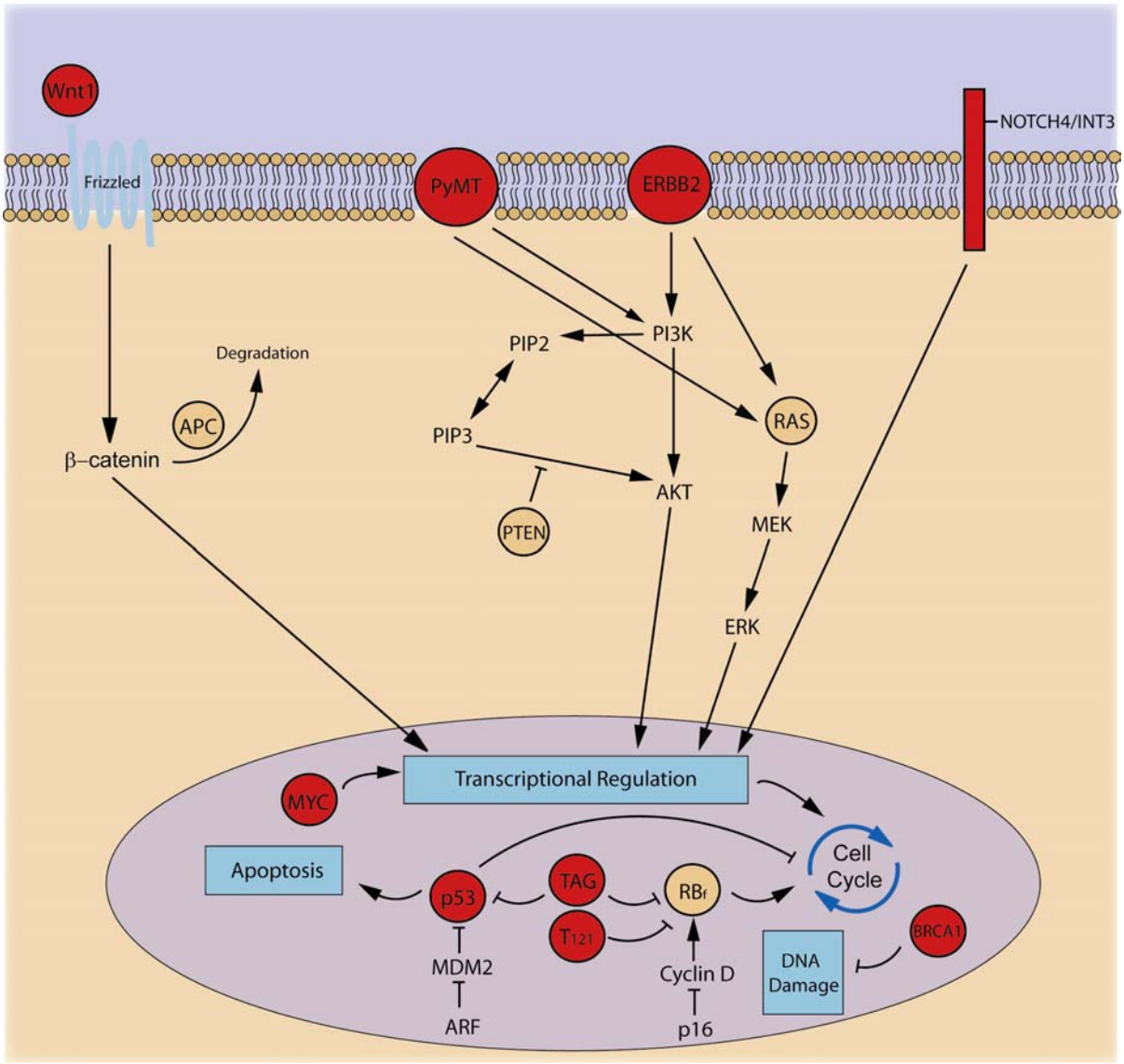


Figure 1.1. Molecular pathways involved in mouse models of human breast cancer. Many of the murine models (red) mimic genetic alterations seen in human breast cancer including inactivation of p53, BRCA1, and Rb, and activation/overexpression of MYC and HER2.

The p53 tumor suppressor is the most commonly altered gene in human breast cancer. It has been found to be mutated in about 30-40% of all human breast tumors with, as described above, much higher frequency in the two ER negative subtypes[3]. Mutations in p53 are also responsible for Li-Fraumeni syndrome, a hereditary disease predisposing patients to a spectrum of tumor types including breast. p53 plays an important role in protecting the genome from DNA damage via blocking the cell cycle in the G1 phase, such that the cell is able to repair genomic damage. If the cell is beyond repair, p53 can induce apoptosis or senescence. When p53 function is removed, aberrant cell growth continues unchecked and the cells become genomically unstable. Mice homozygous or heterozygous for a mutant p53 allele have been shown to present with many different tumor types, but primarily lymphomas. These mice rarely develop mammary tumors. However, it has been shown that modifier genes present in the BALB/c genetic background shifted the spectrum of tumors in p53+/- animals to include a significant increase in mammary gland carcinomas with exposure to ionizing radiation decreasing the latency to tumor formation[39, 40]. Modifier genes on the BALB/c background have recently been mapped[41].

The retinoblastoma 1 (pRb) gene was found to be nonfunctional in up to 25% of human breast cancers. There is also evidence of frequent alterations of other members of the pRb pathway in human breast cancer including p16INK4a loss [42] and cyclin D1 amplification or overexpression [43]. pRb plays an important role in cell cycle progression, when cells exit G0 or G1 and enter S phase. In quiescent or early G1-phase cells, pRb is hypophosphorylated and associates with specific members of the E2F transcription factor family, converting them to active transcriptional repressors. Cell cycle progression from G

to S phase occurs when complexes of D-type cyclins/CDK4/CDK6 phosphorylate pRb, thereby allowing E2Fs to direct transcription of S phase genes. The expression of the early region of Simian virus 40 large and small T-antigen (SV40-T-Ag) has been targeted to the mouse mammary gland using both the C3(1) promoter and the WAP promoter with both models being highly transforming. SV40-T-Ag induces transformation through the inactivation of p53 and the Rb family members. In addition, expression of T121, a fragment of SV40 T antigen that binds to and inactivates Rb family members, using the WAP promoter, also leads to the formation of mammary adenocarcinomas. In this model, heterozygosity for a p53 null allele was shown to significantly shorten tumor latency with most tumors undergoing a subsequent loss of the wild-type allele[44]. These models show that inducing the deficiency of both Rb family members and p53 can induce tumor formation in the mammary gland, which is also seen in the most aggressive human breast tumors.

Mutations in BRCA1 have been identified as a hereditary cause of human breast cancer. Tumors from carriers of BRCA1 mutations were shown to cluster with "basal-like" tumors[3, 45]. BRCA1 has not, however, been found to be somatically mutated in sporadic breast cancer, although methylation of BRCA1 is seen (infrequent in basal-like[46]). BRCA1 has been implicated in the regulation of DNA repair and gene transcription and in the maintenance of genome integrity. Several mouse models have been generated to assess the functions of BRCA1 in the mammary gland. Unlike humans, mice carrying a heterozygous mutation in this gene do not display a detectable increase in tumor formation. However, when combined with mutations in p53, an increase in mammary tumor incidence was observed[47]. In addition, in mice generated to carry a BRCA1-null allele in addition to

a conditional mutation of BRCA1 in mammary epithelial cells, tumor formation occurred after long latency[48]. Loss of p53 was shown to accelerate tumor formation in this model as well, which again mimics the human scenario where BRCA1 mutant tumors tend to be p53 deficient.

Tumors have been generated in mice infected with the mouse mammary tumor virus (MMTV) by insertional mutagenesis and activation of oncogenes. The oncogenes affected include Int3 (Notch 4)[49] and Wnt1. Mammary specific promoters have also been used to drive these transgenes resulting in lesions that resemble those induced by the virus. Acinar tumors of humans are rare but histologically resemble those from the Wnt1 transgene[26]. There is no direct evidence of Notch 4 or Wnt1 being activated in human breast cancer, however, evidence of their involvement is emerging and includes the observed high expression of wnt pathway members in basal-like tumors in our current human data[50].

Research introduction

Mouse models that mimic human disease can be useful for understanding the mechanisms of development and progression of disease. They can also be useful for testing therapeutics and prevention strategies, but only if we understand what type of human disease each model represents. The identification of accurate and useful models is therefore extremely important. Historically, morphologic and architectural criteria have been used to categorize mammary tumors in mice. More recently, the use of molecular markers by immunohistochemistry has added another dimension to the categorization of mammary tumors[51]. The goal of the work presented here was to use gene expression analysis by

microarray to classify mouse models of breast cancer and determine those models that accurately represent human subtypes. Additionally, we have determined gene expression patterns representative of cell types of origin and oncogenic pathways that are seen in common between the two species that have helped inform us about the human disease. Overall, these gene expression profiles will be used to determine which mouse models are useful for particular genetic and pharmacologic studies.

REFERENCES

1. **American Cancer Society** [www.cancer.org]

2. Vargo-Gogola T, Rosen JM: **Modelling breast cancer: one size does not fit all**. *Nat Rev Cancer* 2007, **7**(9):659-672.

3. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc Natl Acad Sci U S A* 2001, **98**(19):10869-10874.

4. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: **Repeated observation of breast tumor subtypes in independent gene expression data sets**. *Proc Natl Acad Sci U S A* 2003, **100**(14):8418-8423.

5. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L *et al*: **The molecular portraits of breast tumors are conserved across microarray platforms**. *BMC Genomics* 2006, **7**(1):96.

6. Lacroix M, Leclercq G: **About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer**. *Mol Cell Endocrinol* 2004, **219**(1-2):1-7.

7. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P *et al*: **Breast cancer molecular subtypes respond differently to preoperative chemotherapy**. *Clin Cancer Res* 2005, **11**(16):5678-5685.

8. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM: **The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes**. *Clin Cancer Res* 2007, **13**(8):2329-2334.

9. Gaedcke J, Traub F, Milde S, Wilkens L, Stan A, Ostertag H, Christgen M, von Wasielewski R, Kreipe HH: **Predominance of the basal type and HER-2/neu type in brain metastasis from breast cancer**. *Mod Pathol* 2007, **20**(8):864-870.

10. Bryan BB, Schnitt SJ, Collins LC: **Ductal carcinoma in situ with basal-like phenotype: a possible precursor to invasive basal-like breast cancer**. *Mod Pathol* 2006, **19**(5):617-621.

11. Dabbs DJ, Chivukula M, Carter G, Bhargava R: **Basal phenotype of ductal carcinoma in situ: recognition and immunohistologic profile**. *Mod Pathol* 2006, **19**(11):1506-1511.

12. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, Tse CK, Nyante S, Millikan RC: **Identification of a basal-like subtype of breast ductal carcinoma in situ**. *Hum Pathol* 2007, **38**(2):197-204.

13. Tang P, Wang X, Schiffhauer L, Wang J, Bourne P, Yang Q, Quinn A, Hajdu SI: **Relationship between nuclear grade of ductal carcinoma in situ and cell origin markers**. *Ann Clin Lab Sci* 2006, **36**(1):16-22.

14. Saal LH, Holm K, Maurer M, Memeo L, Su T, Wang X, Yu JS, Malmstrom PO, Mansukhani M, Enoksson J *et al*: **PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma**. *Cancer Res* 2005, **65**(7):2554-2559.

15. Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Boussard T, Lonning PE, Tibshirani R, Borresen-Dale AL, Pollack JR: **Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer**. *Genes Chromosomes Cancer* 2006, **45**(11):1033-1040.

16. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T *et al*: **Genomic and transcriptional aberrations linked to breast cancer pathophysiologies**. *Cancer Cell* 2006, **10**(6):529-541.

17. Wang ZC, Lin M, Wei LJ, Li C, Miron A, Lodeiro G, Harris L, Ramaswamy S, Tanenbaum DM, Meyerson M *et al*: **Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers**. *Cancer Res* 2004, **64**(1):64-71.

18. Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, Liao X, Iglehart JD, Livingston DM, Ganesan S: **X chromosomal abnormalities in basal-like human breast cancer**. *Cancer Cell* 2006, **9**(2):121-132.

19. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S *et al*: **Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study**. *Jama* 2006, **295**(21):2492-2502.

20. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Smith LV, Labbok MH, Geradts J, Bensen JT, Jackson S *et al*: **Epidemiology of basal-like breast cancer**. *Breast Cancer Res Treat* 2007.

21. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A *et al*: **Differences in risk factors for breast cancer molecular subtypes in a population-based study**. *Cancer Epidemiol Biomarkers Prev* 2007, **16**(3):439-443.

22. Stewart TA, Pattengale PK, Leder P: **Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes**. *Cell* 1984, **38**(3):627-637.

23. Sharpless NE, Depinho RA: **The mighty mouse: genetically engineered mouse models in cancer drug development**. *Nat Rev Drug Discov* 2006, **5**(9):741-754.

24. Rangarajan A, Weinberg RA: **Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice**. *Nat Rev Cancer* 2003, **3**(12):952-959.

25. Cardiff RD, Anver MR, Gusterson BA, Hennighausen L, Jensen RA, Merino MJ, Rehm S, Russo J, Tavassoli FA, Wakefield LM *et al*: **The mammary pathology of genetically engineered mice: the consensus report and recommendations from the Annapolis meeting**. *Oncogene* 2000, **19**(8):968-988.

26. Cardiff RD, Wellings SR: **The comparative pathology of human and mouse mammary glands**. *J Mammary Gland Biol Neoplasia* 1999, **4**(1):105-122.

27. Schoenenberger CA, Andres AC, Groner B, van der Valk M, LeMeur M, Gerlinger P: **Targeted c-myc gene expression in mammary glands of transgenic mice induces mammary tumours with constitutive milk protein gene transcription**. *Embo J* 1988, **7**(1):169-175.

28. van de Vijver MJ, Mooi WJ, Wisman P, Peterse JL, Nusse R: **Immunohistochemical detection of the neu protein in tissue sections of human breast tumors with amplified neu DNA**. *Oncogene* 1988, **2**(2):175-178.

29. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A: **Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer**. *Science* 1989, **244**(4905):707-712.

30. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL: **Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene**. *Science* 1987, **235**(4785):177-182.

31. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: **Genome-wide analysis of DNA copy-number changes using cDNA microarrays**. *Nat Genet* 1999, **23**(1):41-46.

32. Bouchard L, Lamarre L, Tremblay PJ, Jolicoeur P: **Stochastic appearance of mammary tumors in transgenic mice carrying the MMTV/c-neu oncogene**. *Cell* 1989, **57**(6):931-936.

33. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ: **Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease**. *Proc Natl Acad Sci U S A* 1992, **89**(22):10578-10582.

34. Guy CT, Cardiff RD, Muller WJ: **Activated neu induces rapid tumor progression**. *J Biol Chem* 1996, **271**(13):7673-7678.

35. Ursini-Siegel J, Schade B, Cardiff RD, Muller WJ: **Insights from transgenic mouse models of ERBB2-induced breast cancer**. *Nat Rev Cancer* 2007, **7**(5):389-397.

36. Guy CT, Cardiff RD, Muller WJ: **Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease**. *Mol Cell Biol* 1992, **12**(3):954-961.

37. Muthuswamy SK, Siegel PM, Dankort DL, Webster MA, Muller WJ: **Mammary tumors expressing the neu proto-oncogene possess elevated c-Src tyrosine kinase activity**. *Mol Cell Biol* 1994, **14**(1):735-743.

38. Webster MA, Hutchinson JN, Rauh MJ, Muthuswamy SK, Anton M, Tortorice CG, Cardiff RD, Graham FL, Hassell JA, Muller WJ: **Requirement for both Shc and phosphatidylinositol 3' kinase signaling pathways in polyomavirus middle T-mediated mammary tumorigenesis**. *Mol Cell Biol* 1998, **18**(4):2344-2359.

39. Kuperwasser C, Hurlbut GD, Kittrell FS, Dickinson ES, Laucirica R, Medina D, Naber SP, Jerry DJ: **Development of spontaneous mammary tumors in BALB/c p53 heterozygous mice. A model for Li-Fraumeni syndrome**. *Am J Pathol* 2000, **157**(6):2151-2159.

40. Backlund MG, Trasti SL, Backlund DC, Cressman VL, Godfrey V, Koller BH: **Impact of ionizing radiation and genetic background on mammary tumorigenesis in p53-deficient mice**. *Cancer Res* 2001, **61**(17):6577-6582.

41. Koch JG, Gu X, Han Y, El-Naggar AK, Olson MV, Medina D, Jerry DJ, Blackburn AC, Peltz G, Amos CI *et al*: **Mammary tumor modifiers in BALB/cJ mice heterozygous for p53**. *Mamm Genome* 2007, **18**(5):300-309.

42. Geradts J, Wilson PA: **High frequency of aberrant p16(INK4A) expression in human breast cancer**. *Am J Pathol* 1996, **149**(1):15-20.

43. Buckley MF, Sweeney KJ, Hamilton JA, Sini RL, Manning DL, Nicholson RI, deFazio A, Watts CK, Musgrove EA, Sutherland RL: **Expression and amplification of cyclin genes in human breast cancer**. *Oncogene* 1993, **8**(8):2127-2133.

44. Simin K, Wu H, Lu L, Pinkel D, Albertson D, Cardiff RD, Dyke TV: **pRb Inactivation in Mammary Cells Reveals Common Mechanisms for Tumor Initiation and Progression in Divergent Epithelia**. *PLoS Biol* 2004, **2**(2):E22.

45. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT *et al*: **Gene expression profiling predicts clinical outcome of breast cancer**. *Nature* 2002, **415**(6871):530-536.

46. Matros E, Wang ZC, Lodeiro G, Miron A, Iglehart JD, Richardson AL: **BRCA1 promoter methylation in sporadic breast tumors: relationship to gene expression profiles**. *Breast Cancer Res Treat* 2005, **91**(2):179-186.

47. Cressman VL, Backlund DC, Hicks EM, Gowen LC, Godfrey V, Koller BH: **Mammary tumor formation in p53- and BRCA1-deficient mice**. *Cell Growth Differ* 1999, **10**(1):1-10.

48. Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, Hennighausen L, Wynshaw-Boris A, Deng CX: **Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation**. *Nat Genet* 1999, **22**(1):37-43.

49. Gallahan D, Callahan R: **The mouse mammary tumor associated gene INT3 is a unique member of the NOTCH gene family (NOTCH4)**. *Oncogene* 1997, **14**(16):1883-1890.

50. Burkart MF, Wren JD, Herschkowitz JI, Perou CM, Garner HR: **Clustering microarray-derived gene lists through implicit literature relationships**. *Bioinformatics* 2007, **23**(15):1995-2003.

51. Mikaelian I, Blades N, Churchill G, Fancher K, Knowles B, Eppig J, Sundberg J: **Proteotypic classification of spontaneous and transgenic mammary neoplasms**. *Breast Cancer Res* 2004, **6**:R668-R679.

CHAPTER II

IDENTIFICATION OF CONSERVED GENE EXPRESSION FEATURES BETWEEN MURINE MAMMARY CARCINOMA MODELS AND HUMAN BREAST TUMORS

PREFACE

This work was previously published. My role in this project included instigating, organizing, and designing the study. I collected samples, harvested RNA, and performed microarrays and data analysis. I initiated and participated in writing all aspects of the manuscript and figure preparations. Karl Simin also played a significant role in this project by generating and contributing mouse mammary tumors, performing immunofluorescence, analyzing data, and participating in the writing of the manuscript. Victor Weigman assisted greatly with the data analysis. Many researchers at a number of institutions were involved in tumor sample collection from both mouse and human tumors. Terry Van Dyke contributed samples and reagents and helped with the drafting of this manuscript. Charles Perou was the Principle Investigator, conceived the project, and contributed to the writing of the paper.

Jason I. Herschkowitz, Karl Simin, Victor J. Weigman, Igor Mikaelian, Jerry Usary, Zhiyuan Hu, Karen E. Rasmussen, Laundette P. Jones, Shahin Assefnia, Subhashini Chandrasekharan, Michael G. Backlund, Yuzhi Yin, Andrey I. Khramtsov, Roy Bastein, John Quackenbush, Robert I. Glazer, Powel H. Brown, Jeffrey E. Green, Levy Kopelovich, Priscilla A. Furth,

Juan P. Palazzo, Olufunmilayo I. Olopade, Philip S. Bernard, Gary A. Churchill, Terry Van Dyke, and Charles M. Perou (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors Genome Biology 8(5):R76

ABSTRACT

Background

Although numerous mouse models of breast carcinomas have been developed, we do not know the extent to which any faithfully represent clinically significant human phenotypes. To address this need, we characterized mammary tumor gene expression profiles from thirteen different murine models using DNA microarrays and compared the resulting data to that of human breast tumors.

Results

Unsupervised hierarchical clustering analysis showed that six models (TgWAP-*Myc*, TgMMTV-*Neu*, TgMMTV-*PyMT*, TgWAP-*Int3*, TgWAP-*Tag*, and TgC3(1)-*Tag*) yielded tumors with distinctive and homogeneous expression patterns within each strain. However, in each of four other models (TgWAP-T_{121}, TgMMTV-*Wnt1*, $Brca1^{Co/Co}$;TgMMTV-$Cre;p53^{+/-}$ and DMBA-induced), tumors with a variety of histologies and expression profiles developed. In many models, similarities to human breast tumors were recognized including proliferation and human breast tumor subtype signatures. Significantly, tumors of several models displayed characteristics of human basal-like breast tumors including two models with

induced *Brca1* deficiencies. Tumors of other murine models shared features with human luminal tumors.

Conclusions

Many of the defining characteristics of human subtypes were conserved among mouse models. Although no single model recapitulated all the expression features of a given human subtype, these shared expression features provide a common framework for an improved integration of murine mammary tumor models with human breast tumors.

BACKGROUND

Global gene expression analyses of human breast cancers have identified at least three major tumor subtypes and a normal breast tissue group[1]. Two subtypes are ER-negative with poor patient outcomes[2, 3]; one of these two subtypes is defined by the high expression of HER2/ERBB2/NEU (HER2+/ER-) and the other shows characteristics of basal/myoepithelial cells (basal-like). The third major subtype is ER-positive and Keratin 8/18-positive, and designated the "luminal" subtype. This subtype has been subdivided into good outcome "luminal A" tumors and poor outcome "luminal B" tumors[2, 3]. These studies emphasize that human breast cancers are multiple distinct diseases, with each of the major subtypes likely harboring different genetic alterations and responding distinctly to therapy[4, 5]. Further similar investigations may well identify additional subtypes useful in diagnosis and treatment, however, such research would be accelerated if the relevant disease properties could be accurately modeled in experimental animals. Signatures associated with

specific genetic lesions and biologies can be causally assigned in such models, potentially allowing for refinement of human data.

Significant progress in the ability to genetically engineer mice has led to the generation of models that recapitulate many properties of human cancers[6]. Mouse mammary tumor models have been designed to emulate genetic alterations found in human breast cancers including inactivation of TP53, BRCA1, and RB, and overexpression of MYC and HER2/ERBB2/NEU. Such models have been generated through several strategies including transgenic overexpression of oncogenes, expression of dominant interfering proteins, targeted disruption of tumor suppressor genes, and by treatment with chemical carcinogens[7]. While there are many advantages to using the mouse as a surrogate, there are also potential caveats including differences in mammary physiologies and the possibility of unknown species-specific pathway differences. Furthermore, it is not always clear which features of a human cancer are most relevant for disease comparisons (e.g. genetic aberrations, histological features, tumor biology). Genomic profiling provides a tool for comparative cancer analysis and offers a powerful means of cross-species comparison. Recent studies applying microarray technology to human lung, liver, or prostate carcinomas and their respective murine counterparts have reported commonalities[8-10]. In general, each of these studies focused on a single or few mouse models. Here, we used gene expression analysis to classify a large set of mouse mammary tumor models and human breast tumors. The results provide biological insights among and across the mouse models, and comparisons with human data identify biologically and clinically significant shared features.

RESULTS

Murine tumor analysis.

To characterize the diversity of biological phenotypes present within murine mammary carcinoma models, we performed microarray-based gene expression analyses on tumors from 13 different murine models (Table 1) using Agilent microarrays and a common reference design[1]. We performed 122 microarrays consisting of 108 unique mammary tumors and 10 normal mammary gland samples. Using an unsupervised hierarchical cluster analysis of the data (Appendix IIA), murine tumor profiles indicated the presence of gene sets characteristic of endothelial cells, fibroblasts, adipocytes, lymphocytes, and two distinct epithelial cell types (basal/myoepithelial and luminal). Grouping of the murine tumors in this unsupervised cluster showed that some models developed tumors with consistent, model-specific patterns of expression, while other models showed greater diversity and did not necessarily group together. Specifically, the TgWAP-*Myc*, TgMMTV-*Neu*, TgMMTV-*PyMT*, TgWAP-*Int3* (*Notch4*), TgWAP-*Tag* and TgC3(1)-*Tag* tumors had high within-model correlations. In contrast, tumors from the TgWAP-T_{121}, TgMMTV-*Wnt1*, $Brca1^{Co/Co}$;TgMMTV-Cre;$p53^{+/-}$, and DMBA-induced models showed diverse expression patterns. The $p53^{-/-}$ transplant model tended to be homogenous with 4/5 tumors grouping together, while the $Brca1^{+/-};p53^{+/-}$ IR and $p53^{+/-}$ IR models showed somewhat heterogeneous features between tumors; yet, 6/7 $Brca1^{+/-};p53^{+/-}$ IR and 5/7 $p53^{+/-}$ IR were all present within a single dendrogram branch.

Table 2.1. Summary of mouse mammary tumor models.

Tumor Model	Number of Tumors	Specificity of Lesions	Experimental Oncogenic Lesion(s)	Strain	Reference
WAP-myc	13	WAP[1]	cMyc overexpression	FVB	Sandgren et al., 1995
WAP-Int3	7	WAP	Notch4 overexpression	FVB	Gallahan et al., 1996
WAP-T_{121}	5	WAP	pRb,p107,p130 inactivation	B6D2	Simin et al., 2004
WAP-T_{121}	2	WAP	pRb,p107,p130 inactivation	BALB/cJ	Simin et al., 2004
WAP-Tag	5	WAP	SV40 L-T (pRb, p107, p130, p53, p300 inactivation, others); SV40 s-t	C57Bl/6	Husler et al., 1998
C3(1)-Tag	8	C3(1) [2]	SV40 L-T (pRb, p107, p130, p53, p300 inactivation, others); SV40 s-t	FVB	Maroulakou et al., 1994
MMTV-neu	10	MMTV[3]	Unactivated rat Her2	FVB	Guy et al., 1992
MMTV-Wnt1	11	MMTV	Wnt 1 overexpression	FVB	Tsukamoto et al., 1988
MMTV-PyMT	7	MMTV	Py-MT (activation of Src, PI-3' kinase, and Shc)	FVB	Guy et al., 1992
MMTV-Cre;$Brca1^{Co/Co}$; $p53^{+/-}$	10	MMTV	Brca1 truncation mutant; p53 heterozygous null	C57Bl/6	Xu et al., 1999
$p53^{-/-}$ transplanted	5	none	p53 inactivation	BALB/cJ	Jerry et al., 2000
Medroxyprogesterone-DMBA-induced	11	none	Random DMBA-induced	FVB	Yin et al., 2005
$p53^{+/-}$ irradiated	7	none	p53 heterozygous null, random IR induced	BALB/cJ	Backlund et al., 2001
$Brca1^{+/-}$;$p53^{+/-}$ irradiated	7	none	Brca1 and p53 heterozygous null, random IR induced	BALB/cJ	Cressman et al., 1999

[1]-*whey acidic protein* promoter, commonly restricted to lactating mammary gland luminal cells

[2]-C3(1)- 5' flanking region of the C3(1) component of the rat prostate steroid binding protein (PSBP), expressed in mammary ductal cells.

[3]-MMTV- mouse mammary tumor virus promoter, often expressed in virgin mammary gland epithelium, induced with lactation; often expressed at ectopic sites (e.g. Lymphoid cells, salivary gland, others)

Figure 2.1

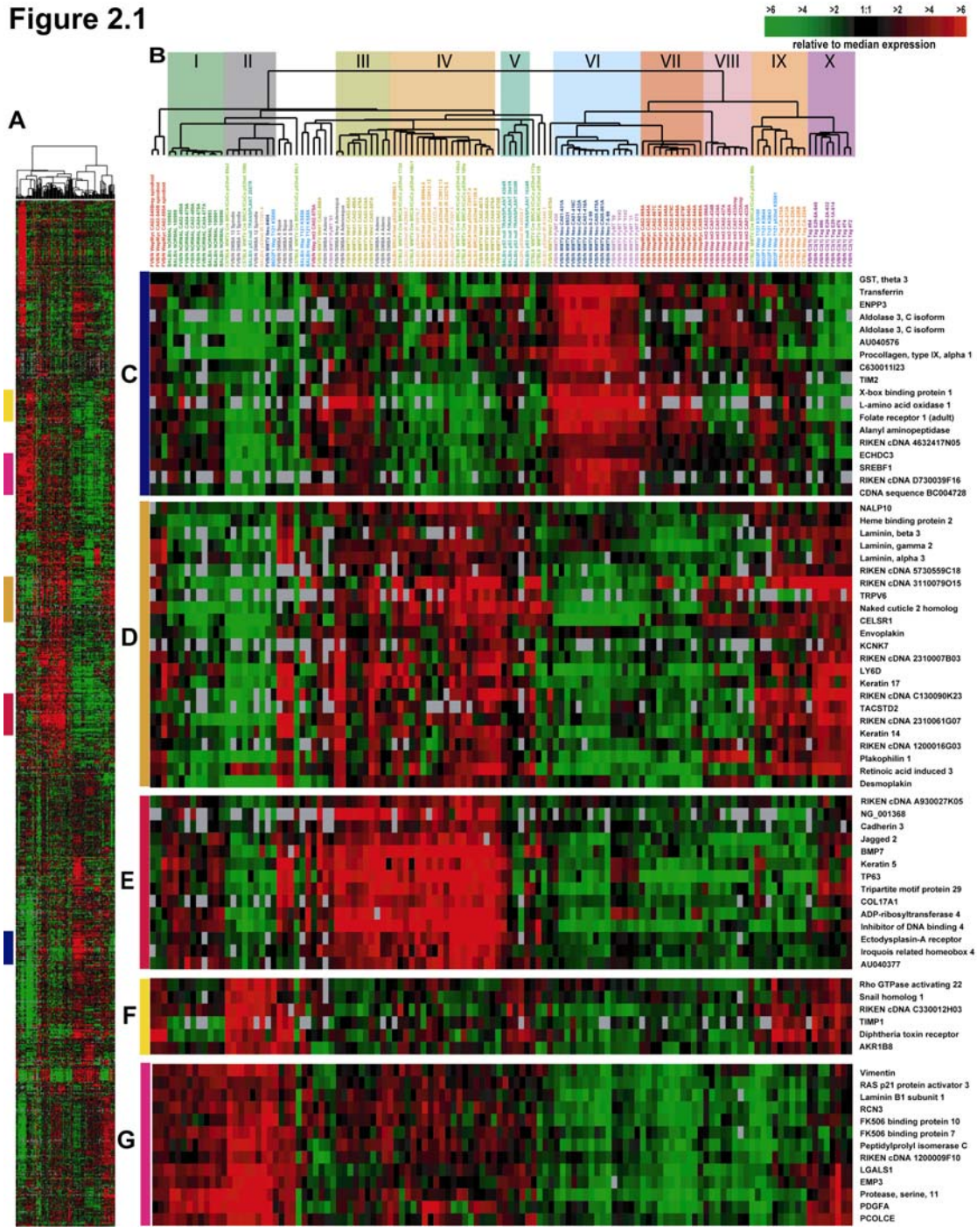

>6 >4 >2 1:1 >2 >4 >6
relative to median expression
A
B
I II III IV V VI VII VIII IX X
C
GST, theta 3
Transferrin
ENPP3
Aldolase 3, C isoform
Aldolase 3, C isoform
AU040576
Procollagen, type IX, alpha 1
C630011I23
TIM2
X-box binding protein 1
L-amino acid oxidase 1
Folate receptor 1 (adult)
Alanyl aminopeptidase
RIKEN cDNA 4632417N05
ECHDC3
SREBF1
RIKEN cDNA D730039F16
CDNA sequence BC004728
D
NALP10
Heme binding protein 2
Laminin, beta 3
Laminin, gamma 2
Laminin, alpha 3
RIKEN cDNA 5730559C18
RIKEN cDNA 3110079O15
TRPV6
Naked cuticle 2 homolog
CELSR1
Envoplakin
KCNK7
RIKEN cDNA 2310007B03
LY6D
Keratin 17
RIKEN cDNA C130090K23
TACSTD2
RIKEN cDNA 2310061G07
Keratin 14
RIKEN cDNA 1200016G03
Plakophilin 1
Retinoic acid induced 3
Desmoplakin
E
RIKEN cDNA A930027K05
NG_001368
Cadherin 3
Jagged 2
BMP7
Keratin 5
TP63
Tripartite motif protein 29
COL17A1
ADP-ribosyltransferase 4
Inhibitor of DNA binding 4
Ectodysplasin-A receptor
Iroquois related homeobox 4
AU040377
F
Rho GTPase activating 22
Snail homolog 1
RIKEN cDNA C330012H03
TIMP1
Diphtheria toxin receptor
AKR1B8
G
Vimentin
RAS p21 protein activator 3
Laminin B1 subunit 1
RCN3
FK506 binding protein 10
FK506 binding protein 7
Peptidylprolyl isomerase C
RIKEN cDNA 1200009F10
LGALS1
EMP3
Protease, serine, 11
PDGFA
PCOLCE

Figure 2.1. Mouse models intrinsic gene set cluster analysis. (A) Overview of the complete 866 gene cluster diagram. (B) Experimental sample associated dendrogram colored to indicate 10 groups. (C) Luminal epithelial gene expression pattern that is highly expressed in TgMMTV-*PyMT*, TgMMTV-*Neu*, and TgWAP-*myc* tumors. (D) Genes encoding components of the basal lamina. (E) A second basal epithelial cluster of genes including *Keratin 5*. (F) Genes expressed in fibroblast cells and implicated in epithelial to mesenchymal transition including *snail homolog 1*. (G) A second mesenchymal cluster that is expressed in normals. See Additional data file 2 for the complete cluster diagram with all gene names.

As with previous human tumor studies[1, 3], we performed an "intrinsic" analysis to select genes consistently representative of groups/classes of murine samples. In the human studies, expression variation for each gene was determined using biological replicates from the same patient, and the "intrinsic genes" identified by the algorithm had relatively low variation within biological replicates and high variation across individuals. In contrast, in this mouse study we applied the algorithm to groups of murine samples defined by an empirically determined correlation threshold of > 0.65 using the dendrogram from Additional data file 2. This "intrinsic" analysis yielded 866 genes that we then used in a hierarchical cluster analysis (Figure 2.1). This analysis identified ten potential groups containing five or more samples each, including a normal mammary gland group (Group I) and nine tumor groups (designated Groups II-X).

In general, these 10 groups were contained within four main categories that included (Figure 2.1B; left to right): A) the normal mammary gland samples (Group I) and tumors with mesenchymal characteristics (Group II), B) tumors with basal/myoepithelial features (Groups III-V), C) tumors with luminal characteristics (Groups VI-VIII), and D) tumors containing mixed characteristics (Groups IX, X). Group I contained all normal mammary gland samples, which showed a high level of similarity regardless of strain, and was characterized by the high expression of basal/myoepithelial (Figure 2.1E) and mesenchymal features including *vimentin* (Figure 2.1G). Group II samples were derived from several models (2/10 $Brca1^{Co/Co}$;TgMMTV-Cre;$p53^{+/-}$, 3/11 DMBA-induced, 1/5 $p53^{-/-}$ transplant, 1/7 $p53^{+/-}$ IR, 1/10 TgMMTV-*Neu* and 1/7 TgWAP-T_{121}) and also showed high expression of mesenchymal features (Figure 2.1G) that were shared with the normal samples in addition to

a second highly expressed mesenchymal-like cluster that contained *snail homolog 1* (a gene implicated in epithelial-mesenchymal transition[11]), the later of which was not expressed in the normal samples (Figure 2.1F). Two TgWAP-*Myc* tumors at the extreme left of the dendrogram, which showed a distinct spindloid histology, also expressed these mesenchymal-like gene features. Further evidence for a mesenchymal phenotype for Group II tumors came from Keratin 8/18 (K8/18) and Smooth Muscle Actin (SMA) immunofluorescence (IF) analyses, which showed that most spindloid tumors were K8/18-negative and SMA-positive (Figure 2.2L).

The second large category contained Groups III-V, with Group III (4/11 DMBA-induced and 5/11 *Wnt1*), Group IV (7/7 Brca1$^{+/-}$;*p53*$^{+/-}$ IR, 4/10 *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/-}$, 4/6 *p53*$^{+/-}$ IR and 3/11 *Wnt1*) and Group V (4/5 *p53*$^{-/-}$ transplant and 1/6 *p53*$^{+/-}$ IR), showing characteristics of basal/myoepithelial cells (Figure 2.1D and E). These features were encompassed within two expression patterns. One cluster included *Keratin 14*, *17* and *LY6D* (Figure 2.1D); *Keratin 17* is a known human basal-like tumor marker[1, 12], while *LY6D* is a member of the Ly6 family of GPI-anchored proteins that is highly expressed in head & neck squamous cell carcinomas[13]. This cluster also contained components of the basement membrane (e.g . *Laminins*) and hemidesmosomes (e.g. *Envoplakin* and *Desmoplakin*), which link the basement membrane to cytoplasmic keratin filaments. A second basal/myoepithelial cluster highly expressed in Group III and IV tumors and a subset of DMBA tumors with squamous morphology, was characterized by high expression of *ID4*, *TRIM29*, and *Keratin 5* (Figure 2.1E), the latter of which is another human basal-like tumor marker[1, 12]. This gene set is expressed in a smaller subset of models compared to the set described above (Figure

2.1D), and is lower or absent in most Group V tumors. As predicted by gene expression data, most of these tumors stained positive for *Keratin 5* (K5) by IF (Figure 2.2G-K).

The third category of tumors (Groups VI-VIII) contained many of the "homogenous" models, all of which showed a potential "luminal" cell phenotype: Group VI contained the majority of the TgMMTV-*Neu* (9/10) and TgMMTV-*PyMT* (6/7) tumors, while Groups VII and VIII contained most of the TgWAP-*Myc* tumors (11/13) and TgWAP-*Int3* samples (6/7), respectively. A distinguishing feature of these tumors (in particular Group VI) was the high expression of *XBP1* (Figure 2.1C), which is a human luminal tumor-defining gene[14-17]. These tumors also expressed tight junction structural component genes including *Occludin*, *Tight Junction Protein 2* and *3*, and the luminal cell *Keratins 8/18* (K8/18) (not shown). IF for K8/18 and K5 confirmed that these tumors all exclusively expressed K8/18 (Fig. 2.2B-F).

Finally, Group IX (1/10 $Brca1^{Co/Co}$;TgMMTV-Cre;$p53^{+/-}$, 4/7 TgWAP-T_{121} tumors and 5/5 TgWAP-*Tag* tumors) and Group X (8/8 TgC3(1)-*Tag*) tumors were present at the far right and showed "mixed" characteristics; in particular, the Group IX tumors showed some expression of luminal (Figure 2.1C), basal (Figure 2.1D) and mesenchymal genes (Figure 2.1F), while Group X tumors expressed basal (Figure 2.1E+F) and mesenchymal genes (Figure 2.1F+G).

Figure 2.2

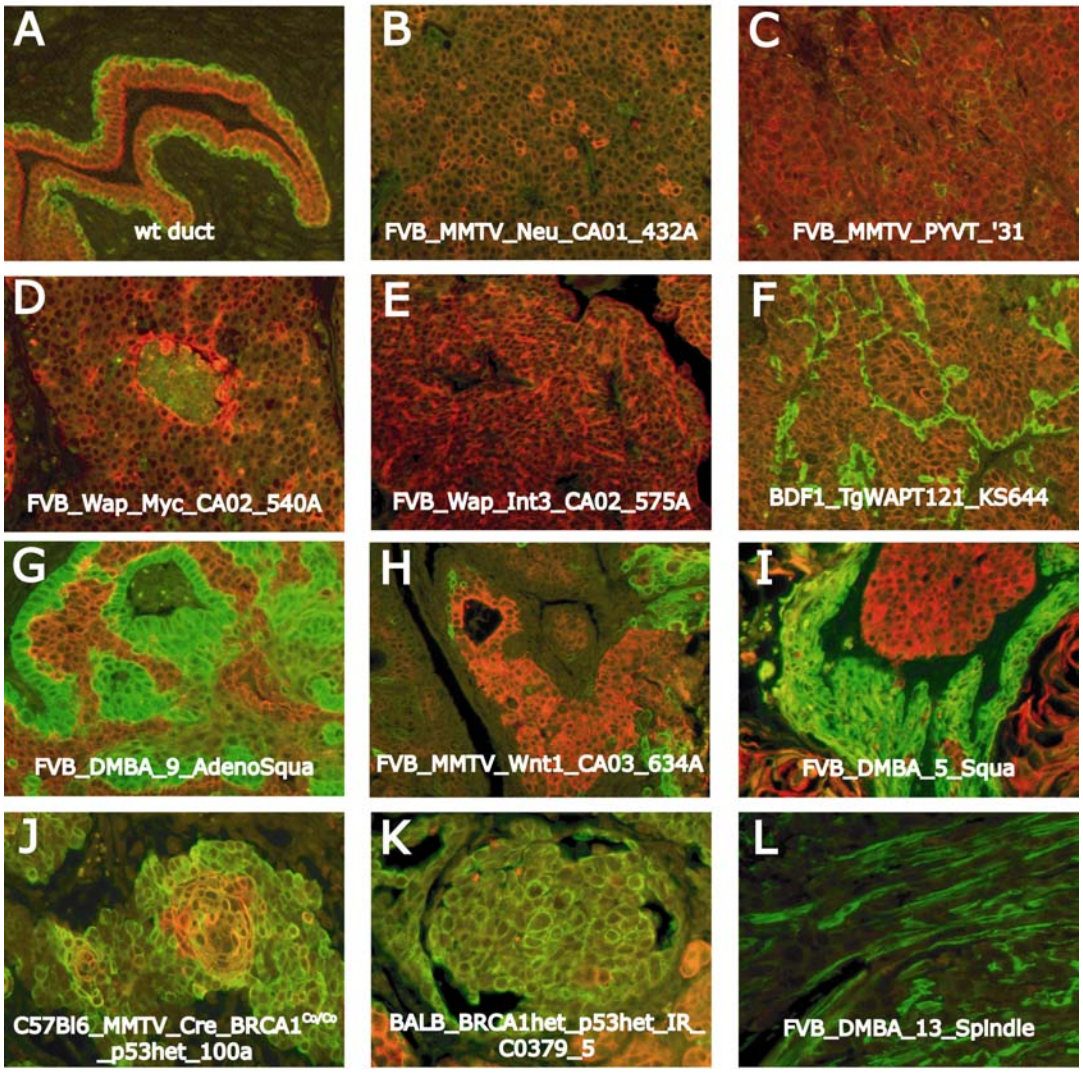

A
wt duct
B
FVB_MMTV_Neu_CA01_432A
C
FVB_MMTV_PYVT_'31
D
FVB_Wap_Myc_CA02_540A
E
FVB_Wap_Int3_CA02_575A
F
BDF1_TgWAPT121_KS644
G
FVB_DMBA_9_AdenoSqua
H
FVB_MMTV_Wnt1_CA03_634A
I
FVB_DMBA_5_Squa
J
C57Bl6_MMTV_Cre_BRCA1Co/Co
_p53het_100a
K
BALB_BRCA1het_p53het_IR_
C0379_5
L
FVB_DMBA_13_Spindle

Figure 2.2. Immunofluorescence staining of mouse samples for basal/myoepithelial and luminal cytokeratins. (A) Wild type mammary gland stained for Keratins 8/18 (red) and Keratin 5 (green) shows K8/18 expression in luminal epithelial cells and K5 expression in basal/myoepithelial cells. (B-F) Mouse models that showed luminal-like gene expression patterns stained with K8/18 (red) and K5 (green). (G-K) Tumor samples that showed basal-like, or mixed luminal and basal characteristics by gene expression, stained for K8/18 (red) and K5 (green). (J) A subset of *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/-}$ tumors showing nodules of K5/K8/18 double positive cells. (L) a splindloid tumor stained for K8/18 (red) and Smooth Muscle Actin (green).

Immunofluorescence analyses showed that, as in humans[12, 18], the murine basal-like models tended to express K5 while the murine luminal models expressed only K8/18. However, some of the murine basal-like models developed tumors that harbored nests of cells of both basal (K5+) and luminal (K8/18+) cell lineages. For example, in some TgMMTV-*Wnt1*[19], DMBA-induced (Figure 2.2G, I), and *Brca1*-deficient strain tumors, distinct regions of single positive K5 and K8/18 cells were observed within the same tumor. Intriguingly, in some $Brca1^{Co/Co}$;TgMMTV-Cre;$p53^{+/-}$ samples, nodules of double-positive K5 and K8/18 cells were identified, suggestive of a potential transition state or precursor/stem cell population (Figure 2.2J), while in some TgMMTV-*Wnt1* (Figure 2.2H) [19]and *Brca1*-deficient tumors, large regions of epithelioid cells were present that had little to no detectable K5 or K8/18 staining (data not shown).

The reproducibility of these groups was evaluated using "Consensus Clustering (CC)"[20]. CC using the intrinsic gene list showed strong concordance with the results from Figure 2.1 and supports the existence of most of the groups identified using hierarchical clustering analysis (Appendix IIB). However, our further division of some of the CC-defined groups appears justified based upon biologic knowledge. For instance, hierarchical clustering separated the normal mammary gland samples (Group I) and the histologically distinct spindloid tumors (Group II), which were combined into a single group by CC. Groups VI (TgMMTV-*Neu* and *PyMT*) and VII (TgWAP-*Myc*) were likewise separated by hierarchical clustering, but CC placed them into a single category. CC was also performed using all genes that were expressed and varied in expression, which showed far less concordance with the intrinsic list based classifications, and which often separated tumors from individual models

into different groups (Appendix IIB C, bottom most panel); for example, the TgMMTV-*Neu* tumors were separated into 2-3 different groups, whereas these were distinct and single groups when analyzed using the intrinsic list. This is likely due to the presence or absence of gene expression patterns coming from other cell types (*i.e.* lymphocytes, fibroblasts, *etc*.) in the "all genes" list, which causes tumors to be grouped based upon qualities not coming from the tumor cells[1].

Mouse-Human combined unsupervised analysis.

The murine gene clusters were reminiscent of gene clusters identified previously in human breast tumor samples. To more directly evaluate these potential shared characteristics, we performed an integrated analysis of the mouse data presented here with an expanded version of our previously reported human breast tumor data. The human data were derived from 232 microarrays representing 184 primary breast tumors and 9 normal breast samples also assayed on Agilent microarrays and using a common reference strategy (combined human data sets of [21-23] plus 58 new patients/arrays). To combine the human and mouse data sets, we first used the Mouse Genome Informatics database to identify well-annotated mouse and human orthologous genes. We then performed a Distance Weighted Discrimination correction, which is a supervised analysis method that identifies systematic differences present between two data sets and makes a global correction to compensate for these global biases[24]. Finally, we created an unsupervised hierarchical cluster of the mouse and human combined data (Figure 2.3).

This analysis identified many shared features including clusters that resemble the cell-lineage clusters described above. Specifically, human basal-like tumors and murine *Brca1*$^{+/-}$;*p53*$^{+/-}$;IR, *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/-}$, TgMMTV-*Wnt1*, and some DMBA-induced tumors were characterized by the high expression of *Laminin gamma 2*, *Keratins 5*, *6B*, *13*, *14*, *15*, *TRIM29*, *c-KIT* and *CRYAB* (Figure 2.3B), the last of which is a human basal-like tumor marker possibly involved in resistance to chemotherapy[25]. As described above, the *Brca1*$^{+/-}$;*p53*$^{+/-}$;IR, some *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/}$, DMBA-induced, and TgMMTV-*Wnt1* tumors stained positive for K5 by IF, and human basal-like tumors tend to stain positive using a K5/6 antibody[1, 12, 18, 26], thus showing that basal-like tumors from both species share K5 protein expression as a distinguishing feature.

The murine and human "luminal tumor" shared profile was not as similar as the shared basal profile, but did include the high expression of *SPDEF*, *XBP1* and *GATA3* (Figure 2.3C), and both species luminal tumors also stained positive for K8/18 (Figure 2.2 and see [18]). For many genes in this luminal cluster, however, the relative level of expression differed between the two species. For example, some genes were consistently high across both species' tumors (*e.g. XBP1, SPDEF* and *GATA3*), while others including *TFF, SLC39A6,* and *FOXA1*, were high in human luminal tumors and showed lower expression in murine tumors. Of note is that the human luminal epithelial gene cluster always contains the *Estrogen-Receptor* (*ER*) and many estrogen-regulated genes including *TFF1* and *SLC39A6*[22]; since most murine mammary tumors, including those profiled here, are ER-negative, the apparent lack of involvement of ER and most ER-regulated genes could explain

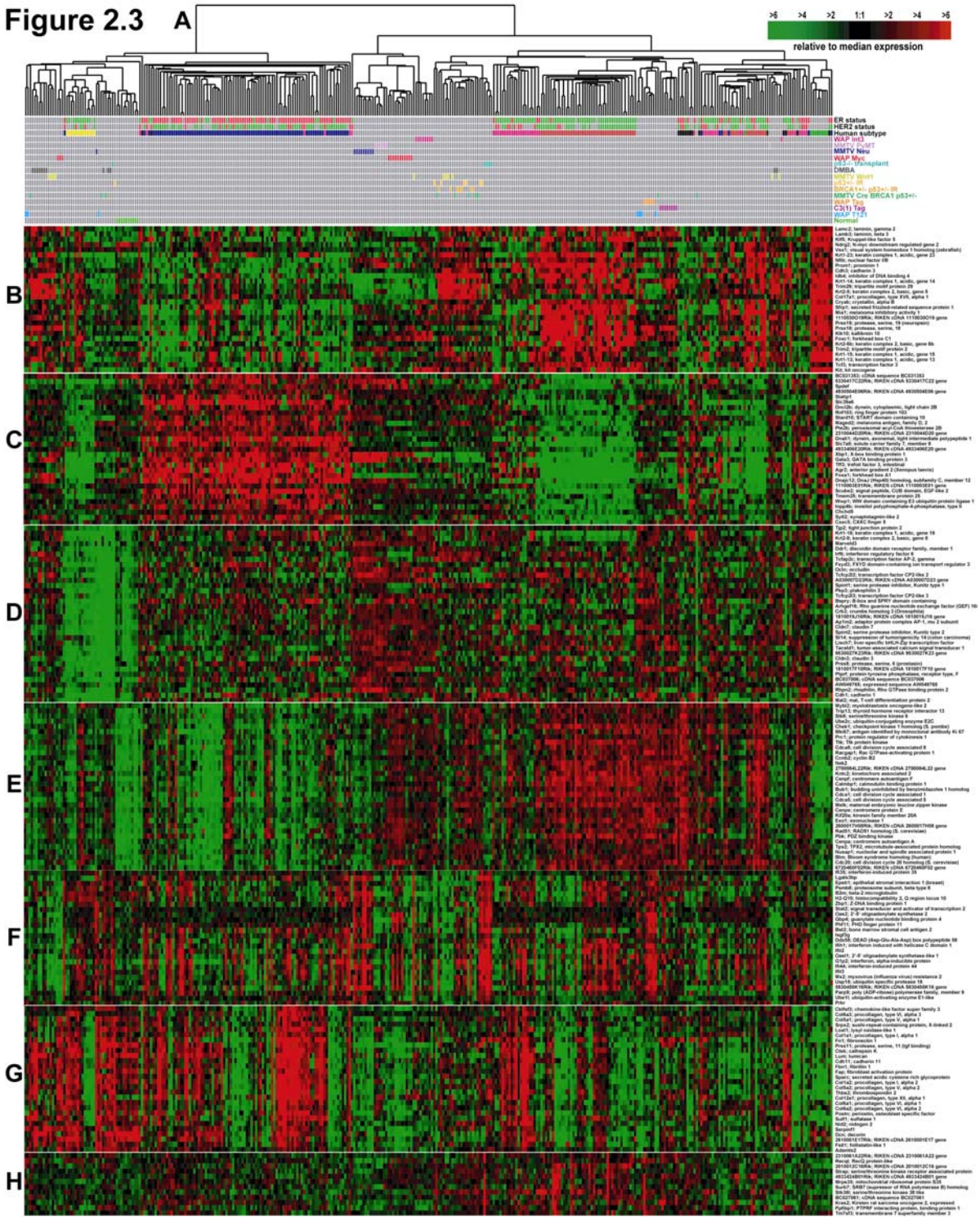


Figure 2.3
A
>6
>4
>2
1:1
>2
>4
>6
relative to median expression
ER status
HER2 status
Human subtype
WAP Int3
MMTV PyMT
MMTV Neu
WAP Myc
p53+/- transplant
DMBA
MMTV Wnt1
p53+/- IR
BRCA1+/- p53+/- IR
MMTV Cre BRCA1 p53+/-
WAP Tag
C3(1) Tag
WAP T121
Normal
B
C
D
E
F
G
H

Figure 2.3. Unsupervised cluster analysis of the combined gene expression data for 232 human breast tumor samples and 122 mouse mammary tumor samples. (A) A color-coded matrix below the dendrogram identifies each sample; the first two rows show clinical ER and HER2 status respectively with red=positive, green=negative, and gray=not tested; the third row includes all human samples colored by intrinsic subtype as determined from Additional data file 6; red=basal-like, blue=luminal, pink= HER2+/ER-, yellow=claudin-low and green=normal breast-like. The remaining rows correspond to murine models indicated at the right. (B) A gene cluster containing basal epithelial genes. (C) A luminal epithelial gene cluster that includes *XBP1* and *GATA3*. (D) A second luminal cluster containing *Keratins 8* and *18*. (E) Proliferation gene cluster. (F) Interferon-regulated genes. (G) Fibroblast/mesenchymal enriched gene cluster. (H) The *Kras2* amplicon cluster. See Additional data file 5 for the complete cluster diagram.

the difference in expression for some of the human luminal epithelial genes that show discordant expression in mice.

Several other prominent and noteworthy features were also identified across species including a "proliferation" signature that includes the well documented proliferation marker Ki-67 (Figure 2.3E)[1, 27, 28] and an interferon-regulated pattern (Figure 2.3F)[27]. The proliferation signature was highest in human basal-like tumors and in the murine models with impaired pRb function (*i.e.* Group IX and X tumors). Currently, the growth regulatory impact of interferon-signaling in human breast tumors is not understood, and murine models that share this expression feature (TgMMTV-*Neu*, TgWAP-Tag, *p53*$^{-/-}$ transplants, and spindloid tumors) may provide a model for future studies of this pathway. A fibroblast profile (Figure 2.3G) that was highly expressed in murine samples with spindloid morphology and in the TgWAP-*Myc* "spindloid" tumors was also observed in many human luminal and basal-like tumors; however, on average, this profile was expressed at lower levels in the murine tumors, which is consistent with the relative epithelial to stromal cell proportions seen histologically.

Through these analyses we also discovered a potential new human subtype (Figure 2.3, top line-yellow group). This subtype that was apparent in both the human only and mouse-human combined data set, is referred to as the "claudin-low" subtype and is characterized by the low expression of genes involved in tight junctions and cell-cell adhesion including *Claudins 3,4, 7*, *Occludin*, and *E-cadherin* (Figure 2.3D). These human tumors (n=13) also showed low expression of luminal genes, inconsistent basal gene expression, and high expression of lymphocyte and endothelial cell markers. All but one

tumor in this group was clinically ER-negative, and all were diagnosed as Grade II or III infiltrating ductal carcinomas (Appendix IIC for representative H&E images); thus, these tumors do not appear to be lobular carcinomas as might be predicted by their low expression of *E-cadherin*. The uniqueness of this group was supported by shared mesenchymal expression features with the murine spindloid tumors (Figure 2.3G), which cluster near these human tumors and also lack expression of the *Claudin* gene cluster (Figure 2.3D). Further analyses will be required to determine the cellular origins of these human tumors.

A common region of amplification across species

The murine C3(1)-*Tag* tumors and a subset of human basal-like tumors showed high expression of a cluster of genes including *Kras2*, *Ipo8*, *Ppfibp1*, *Surb*, and *Cmas* that are all located in a syntenic region corresponding to human chromosome 12p12 and mouse chromosome 6 (Figure 2.3H). *Kras2* amplification is associated with tumor progression in the C3(1)-*Tag* model[29], and haplo-insufficiency of *Kras2* delays tumor progression[30]. High co-expression of *Kras2*-linked genes prompted us to test whether DNA copy number changes might also account for the high expression of *Kras2* among a subset of the human tumors. Indeed, 9 of 16 human basal-like tumors tested by qPCR had increased genomic DNA copy numbers at the *KRAS2* locus; however, no mutations were detected in *KRAS2* in any of these 16 basal-like tumors. In addition, Van Beers *et al*. reported that this region of human Chromosome 12 is amplified in 47% of *BRCA1*-associated tumors by comparative genomic hybridization (CGH) analysis[31]; *BRCA1*-associated tumors are known to exhibit a basal-like molecular profile[3, 32]. In cultured human mammary epithelial cells (HMEC), which show basal/myoepithelial characteristics[1, 33], both high oncogenic H-ras and SV40

Large T-antigen expression are necessary for transformation[34]. Taken together, these findings suggest that amplification of *KRAS2* may either influence the cellular phenotype or define a susceptible target cell type for basal-like tumors.

Mouse-Human shared intrinsic features.

To simultaneously classify mouse and human tumors, we identified the gene set that was in common between a human breast tumor intrinsic list (1300 genes described in Hu et al.[21]) and the mouse intrinsic list developed here (866 genes). The overlap of these two lists totaled 106 genes, which when used in a hierarchical clustering analysis (Figure 2.4) identifies four main groups: the leftmost group contained all the human basal-like, "claudin-low", and 5/44 HER2+/ER- tumors, and the murine C3(1)-*Tag,* TgWAP-*Tag*, and spindloid tumors. The second group (left to right) contained the normal samples from both humans and mice, a small subset of human 6/44 HER2+/ER- and 10/92 luminal tumors, and a significant portion of the remaining murine basal-like models. By clinical criteria, nearly all human tumors in these two groups were clinically classified as ER-negative.

The third group contained 33/44 human HER2+/ER- tumors and the murine TgMMTV-*Neu*, MMTV-PyMT and TgWAP-*Myc* samples. Although the human HER2+/ER- tumors are predominantly ER-negative, this comparative genomic analysis and their keratin expression profiles as assessed by IHC, suggests that the HER2+/ER- human tumors are "luminal" in origin as opposed to showing basal-like features[18]. The fourth and right-most group was composed of ER-positive human luminal tumors and lastly, the mouse TgWAP-

Figure 2.4

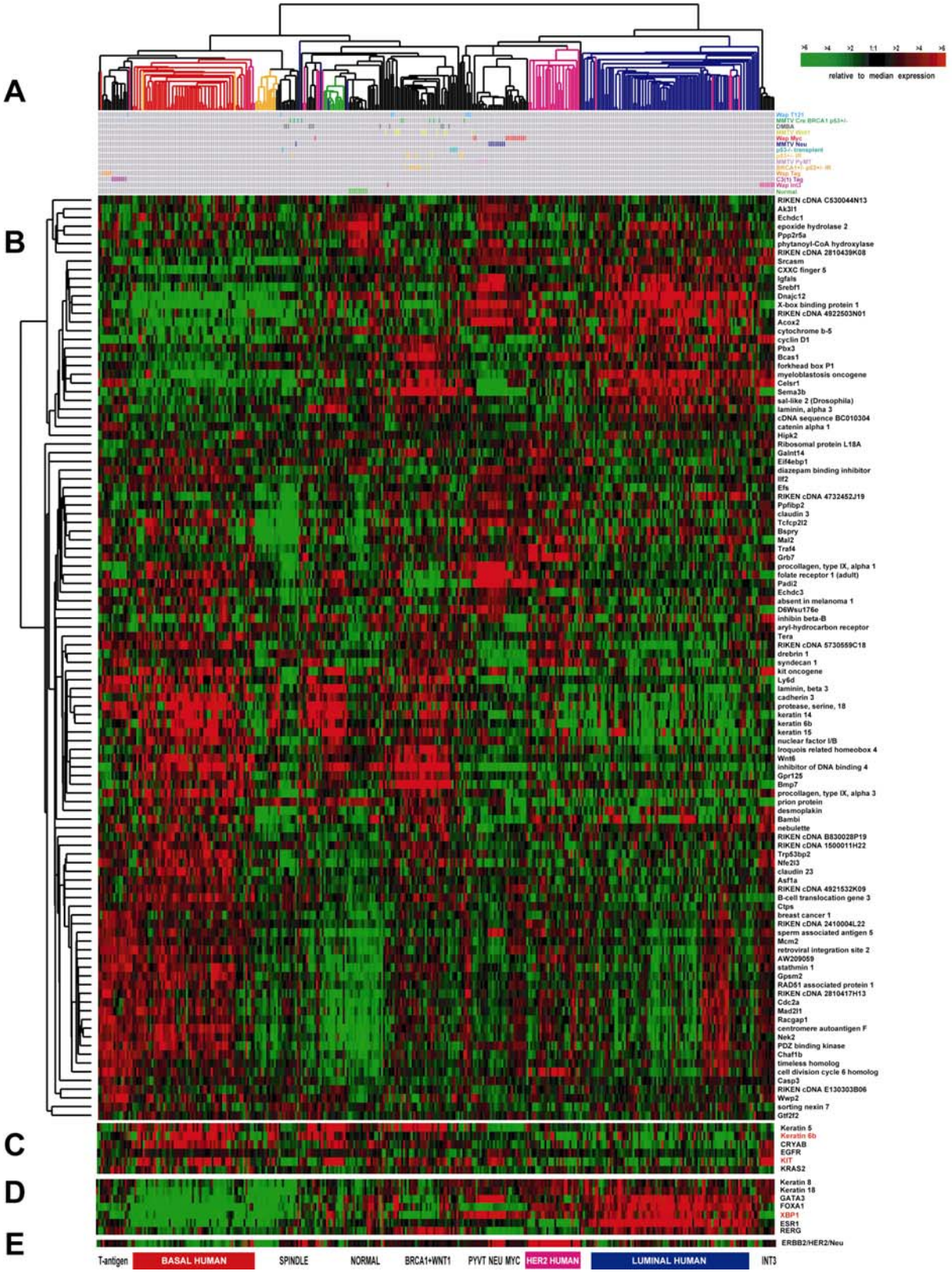

A
B
C
D
E
relative to median expression
Wap T121
MMTV Cre BRCA1 p53+/-
DMBA
MMTV PyVT
Wap Myc
MMTV Neu
p53+/- transplant
p53-/-
MMTV PyVT
BRCA1+/- p53+/- IR
Wap Tag
C3(1) Tag
Wap INT
Normal
RIKEN cDNA C530044N13
Ak3l1
Echdc1
epoxide hydrolase 2
Ppp2r5a
phytanoyl-CoA hydroxylase
RIKEN cDNA 2810439K08
Srcasm
CXXC finger 5
Igfals
Srebf1
Dnajc12
X-box binding protein 1
RIKEN cDNA 4922503N01
Acox2
cytochrome b-5
cyclin D1
Pbx3
Bcas1
forkhead box P1
myeloblastosis oncogene
Celsr1
Sema3b
sal-like 2 (Drosophila)
laminin, alpha 3
cDNA sequence BC010304
catenin alpha 1
Hipk2
Ribosomal protein L18A
Galnt14
Eif4ebp1
diazepam binding inhibitor
Iif2
Efs
RIKEN cDNA 4732452J19
Ppfibp2
claudin 3
Tcfcp2l2
Bspry
Mal2
Traf4
Grb7
procollagen, type IX, alpha 1
folate receptor 1 (adult)
Padi2
Echdc3
absent in melanoma 1
D6Wsu176e
inhibin beta-B
aryl-hydrocarbon receptor
Tera
RIKEN cDNA 5730559C18
drebrin 1
syndecan 1
kit oncogene
Ly6d
laminin, beta 3
cadherin 3
protease, serine, 18
keratin 14
keratin 6b
keratin 15
nuclear factor I/B
Iroquois related homeobox 4
Wnt6
inhibitor of DNA binding 4
Gpr125
Bmp7
procollagen, type IX, alpha 3
prion protein
desmoplakin
Bambi
nebulette
RIKEN cDNA B830028P19
RIKEN cDNA 1500011H22
Trp53bp2
Nfe2l3
claudin 23
Asf1a
RIKEN cDNA 4921532K09
B-cell translocation gene 3
Ctps
breast cancer 1
RIKEN cDNA 2410004L22
sperm associated antigen 5
Mcm2
retroviral integration site 2
AW209069
stathmin 1
Gpsm2
RAD51 associated protein 1
RIKEN cDNA 2810417H13
Cdc2a
Mad2l1
Racgap1
centromere autoantigen F
Nek2
PDZ binding kinase
Chaf1b
timeless homolog
cell division cycle 6 homolog
Casp3
RIKEN cDNA E130303B06
Wwp2
sorting nexin 7
Gtf2f2
Keratin 5
Keratin 6b
CRYAB
EGFR
KIT
KRAS2
Keratin 8
Keratin 18
GATA3
FOXA1
XBP1
ESR1
RERG
ERBB2/HER2/Neu
T-antigen
BASAL HUMAN
SPINDLE
NORMAL
BRCA1+WNT1
PYVT NEU MYC
HER2 HUMAN
LUMINAL HUMAN
INT3

Figure 2.4. Cluster analysis of mouse and human tumors using the subset of genes common to both species intrinsic lists (106 total genes). (A) Experimental sample associated dendrogram color coded according to human tumor subtype and with a matrix below showing murine tumor origins. (B) The complete 106 gene cluster diagram. (C) Close-up of genes known to be important for human basal-like tumors. (D) Close-up of genes known to be important for human luminal tumors including ER. (E) Expression pattern of HER2/ERBB2/NEU.

Int3 (*Notch4*) tumors were in a group by themselves. This data shows that although many mouse and human tumors were located on a large dendrogram branch that contained most murine luminal models and human HER2+/ER- tumors, none of the murine models we tested showed a strong human "luminal " phenotype that is characterized by the high expression of *ER*, *GATA3*, *XBP1* and *FOXA1*. These analyses suggest that the murine luminal models like MMTV-*Neu,* showed their own unique profile that was a relatively weak human luminal phenotype that is missing the ER-signature. Presented at the bottom of Figure 2.4 are biologically important genes discussed here, genes previously shown to be human basal-like tumor markers (Figure 2.4C), human luminal tumor markers including ER (Figure 2.4D), and HER2/ERBB2/NEU (Figure 2.4E).

A comparison of gene sets defining human tumors and murine models.

We used a second analysis method called Gene Set Enrichment Analysis (GSEA)[35], to search for shared relationships between human tumor subtypes and murine models. For this analysis, we first performed a two-class unpaired Significance Analysis of Microarray (SAM)[36] analysis for each of the ten murine groups defined in Figure 2.1, and obtained a list of highly expressed genes that defined each group. Next, we performed similar analyses using each human subtype versus all other human tumors. Lastly, the murine lists were compared to each human subtype list using GSEA, which utilizes both gene list overlap and gene rank (Table 2.2). We found that the murine Groups IX ($p=0.004$) and X ($p=0.001$), which were comprised of tumors from pRb-deficient/p53-deficient models, shared significant overlap with the human basal-like subtype and tended to be anti-correlated with human luminal tumors ($p=0.083$ and 0.006 respectively). Group III murine tumors (TgMMTV-*Wnt1*

mostly) significantly overlapped human normal breast samples (p=0.008), possibly due to the expression of both luminal and basal/myoepithelial gene clusters in both groups. Group IV (*Brca1-deficient* and *Wnt1*) showed a significant association (p=0.058) with the human basal-like profile. The murine Group VI (TgMMTV-*Neu* and TgMMTV-*PyMT*) showed a near significant association (p=0.078) with the human luminal profile and were anti-correlated with the human basal-like subtype (p=0.04). Finally, the murine Group II spindloid tumors showed significant overlap with human "claudin-low" tumors (p=0.001), which further suggests that this may be a distinct and novel human tumor subtype.

We also performed a two-class unpaired SAM analysis using each mouse model as a representative of a pathway perturbation using the transgenic "event" as a means of defining groups. Models that yielded a significant gene list (FDR = 1%) were compared to each human subtype as described above (Appendix IID). The models based upon SV40 T-antigen (all C3(1)-Tag and WAP-Tag tumors) shared significant overlap with the human basal-like tumors (p=0.002) and were marginally anti-correlated with the human luminal class. The BRCA1 deficient models (all Brca1$^{+/-}$;*p53*$^{+/-}$ IR and *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/-}$ tumors) were marginally significant with human basal-like tumors (p=0.088). The TgMMTV-*Neu* tumors were nominally significant (before correction for multiple comparisons) with human luminal tumors (p=0.006) and anti-correlated with human basal-like tumors (p=0.027).

Table 2,2. Gene Set Enrichment Analysis (GSEA) of the 10 murine groups versus 5 human subtypes. Statistically significant findings are highlighted in bold.

Is Class

		BASAL-like		**LUMINAL**		**HER2+/ER-**		**NORMAL**		**claudin-LOW**	
Mouse Class	# genes	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val
I	1882	-	-	0.4625	0.8755	0.5388	0.9137	0.1659	0.5628	0.0048	0.1028
II	912	-	-	-	-	0.5867	0.9609	-	-	0.0021	**0.001**
III	143	0.5289	0.9048	-	-	0.5285	0.9047	0	**0.008**	-	-
IV	1019	0	**0.0581**	-	-	-	-	-	-	-	-
V	34	-	-	0.8492	0.998	0.9324	0.999	-	-	0.0427	0.09274
VI	820	-	-	0.0062	0.0783	0.3536	0.7864	0.8653	0.9769	-	-
VII	851	0.1258	0.3768	-	-	0.5616	0.9137	-	-	-	-
VIII	236	0.1449	0.6098	0.3483	0.8205	-	-	0.01878	0.2349	-	-
IX	462	0.0019	**0.004**	-	-	0.56	0.9509	-	-	-	-
X	338	0	**0.001**	-	-	0.9275	0.998	-	-	-	-

Is Not Class

		BASAL-like		**LUMINAL**		**HER2+/ER-**		**NORMAL**		**claudin-LOW**	
Mouse Class	# genes	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val
I	1882	0.0128	0.1662	-	-	-	-	-	-	-	-
II	912	0.3996	0.8348	0.8601	0.999	-	-	0.3602	0.7655	-	-
III	143	-	-	0.3178	0.7259	-	-	-	-	0.7628	0.991
IV	1019	-	-	0.1833	0.6516	0.398	0.8427	0.2241	0.7255	0.1453	0.6116
V	34	0.86	1	-	-	-	-	0.0656	0.1653	-	-
VI	820	0	**0.04**	-	-	-	-	-	-	0.1043	0.4444
VII	851	-	-	0.1733	0.5151	-	-	0.5403	0.9128	0.1628	0.5215
VIII	236	-	-	-	-	0.1131	0.5305	-	-	0.6427	0.961
IX	462	-	-	0.04305	0.0833	-	-	0.022	**0.037**	0.2612	0.5936
X	338	-	-	0.02236	0.0682	-	-	0.1313	0.3717	0.5437	0.9489

The two most important human breast tumor biomarkers are ER and HER2, therefore, we also analyzed these data relative to these two markers. Of the 232 human tumors assayed here, 137 had ER and HER2 data assessed by IHC and microarray data. As has been noted before[3, 18, 21], there is a very high correlation between tumor intrinsic subtype and ER and HER2 clinical status ($p<0.0001$): for example, 81% of ER+ tumors were of the luminal phenotype, 63% of HER2+ tumors were classified as HER2+/ER-, and 80% of ER- and HER2- tumors were of the Basal-like subtype. Using GSEA, we compared the 10 mouse

classes as defined in Figure 2.1 (Appendix IIE) and the mouse models based gene lists (Appendix IIF) to the human data/gene lists that were obtained by performing supervised analyses based upon human ER and HER2 status (please note that analyses using HER2 status alone (*i.e.* HER2+ vs. HER2-), and ER+ and HER2+ vs. others were not included as human classes because HER2 status alone only yielded genes on the HER2 amplicon, and the ER+ and HER2+ classification did not yield a significant gene list). We found that the murine Groups IX (p=0.009) and X (p=0.003) tumors shared significant overlap with ER- HER2- human tumors and were significantly anti-correlated with human ER+ tumors (p=0.024 and 0.043 respectively). Group VI murine samples (TgMMTV-*Neu* and TgMMTV-*PyMT*) likewise showed the same trend of enrichment with ER+ human tumors and anti-correlation with the ER- HER2- class. Although not perfect, these GSEA results are consistent with our observations from Figures 2.1 and 2.3 and again demonstrate that the basal-like profile is robustly shared between humans and mice, while the luminal profile shows some shared and some distinct features across species.

DISCUSSION

Gene expression profiling of murine tumors and their comparison to human tumors identified characteristics relevant to individual murine models, to murine models in general, and to cancers of both species. First was the discovery that some murine models developed highly similar tumors within models, while others showed heterogeneity in expression and histological phenotypes. For the homogenous models, the study of progression or response to therapy is simplified because confounding variation across individuals is low. An example of this consistency even extended to secondary events that occurred within the TgC3(1)-*Tag*

model, where many tumors shared the amplification and high expression of *Kras2* (Figure 2.3H) - a feature also evident in a subset of human basal-like tumors.

In contrast to the "homogenous" models are models such as TgWAP-T_{121}, DMBA-induced and *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/-}$, where individual tumors within a given model often showed different gene expression profiles and histologies. It is likely that these models fall into one of three scenarios that could explain their heterogeneity: the first represented by the TgWAP-T_{121} model[37], is that the transgene is responsible only for initiating tumorigenesis, leaving progression events to evolve stochastically and with longer latency periods. Such a model would likely give rise to different tumor subtypes depending on the subsequent pathways that are disrupted during tumor progression. A second possibility is that the initiating event generates genomic instability such that multiple distinct pathways can be affected by the experimental causal event, which may be the mechanism in the *Brca1*-inactivation tumors. The third scenario is that the target cell of transformation is a multi-potent progenitor with the ability to undergo differentiation into multiple epithelial lineages, or even mesenchymal lineages (e.g. DMBA-induced and *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/-}$); support for this hypothesis comes from Keratin IF analyses in which, even within a histologically homogenous tumor, two types of epithelial cells are present (Figures 2.2G-K). The presence of subsets of individual cells positive for markers of two epithelial cell types also supports this possibility (Figure 2.2J). Alternative hypotheses include the possibility that multiple cell types sustain transforming events, and also that extensive non-cell-autonomous tissue responses occur. Regardless of the paradigm of transformation for these heterogeneous

models, the study of progression or therapeutic response will best be accomplished by first sub-setting by subtype, and then focusing on biological phenotypes.

There are at least two major applications for genomic comparisons between human tumors and their potential murine counterparts. First, such studies should identify those models that contain individual and/or global characteristics of a particular class of human tumors. Examples of important global characteristics identified here include the classification of murine and human tumors into basal and luminal groups. It appears as if four murine models developed potential luminal-like tumors (TgMMTV-*Neu*, TgMMTV-*PyMT*, TgWAP-*Myc*, and TgWAP-*Int3*), which is not surprising since both MMTV and WAP are thought to direct expression in differentiated alveolar/luminal cells[38, 39]; however, it should be noted that the luminal profile across species was not statistically significant likely due to the lack of ER and ER-regulated genes in the murine luminal tumors. Several murine models did show expression features consistent with human basal-like tumors including the TgC3(1)-*Tag*, TgWAP-*Tag* and *Brca1*-deficient models. The SV40 T-antigen used in the TgC3(1)-*Tag* and TgWAP-*Tag* models inactivates p53 and RB, which also appear to be two likely events that occur in human basal-like tumors because these tumors are known to harbor *p53* mutations[2], have high mitotic grade and the highest expression of proliferation genes (Figure 2.3 [2, 3]), which are known E2F targets[40]. The proliferation signature in human breast cancers, is itself prognostic[41], and is also predictive of response to chemotherapy[42]. These data suggest that human basal-like tumors might have impairment of RB function and highlight an important shared feature of murine and human mammary carcinomas.

The finding that *Brca1* loss (coincident with *p53* mutation), in mice gives rise to tumors with a basal-like phenotype is notable because humans carrying *BRCA1* germline mutations also develop basal-like tumors[3, 32], and most human *BRCA1* mutant tumors are p53-deficient[43, 44]. These data suggest a conserved predisposition of the basal-like cell type, or its progenitor cell, to transform as a result of *BRCA1*, *TP53*, and *RB*-pathway loss. Most DMBA-induced carcinomas also showed basal-like cell lineage features, suggesting that this cell type is also susceptible to DMBA-mediated tumorigenesis. Finally, some TgMMTV-*Wnt1* tumors showed a combination of basal-like and luminal characteristics by gene expression, which is consistent with the observation that tumors of this model generally contain cells from both mammary epithelial lineages[45].

The second major purpose of comparative studies is to determine the extent to which analyses of murine models can inform the human disease and guide further discovery. An example of murine models informing the human disease is encompassed by the analysis of the new potential human subtype discovered here (i.e. claudin-low subtype). Further analysis will be necessary to confirm whether this is a *bona fide* subtype, however, the statistically significant gene overlap with a histologically distinct subset of murine tumors suggests it is a distinct biological entity. A second example of the murine tumors guiding discovery in humans was the common association of a K-Ras containing amplicon in a subset of human basal-like tumors and in the murine basal-like TgC3(1)-*Tag* strain tumors.

An important caveat to all comparative studies is that there are clear biological differences between mice and humans, which may or may not directly impact disease

mechanisms. A potential example of inherent species difference could be the aforementioned biology associated with ER and its downstream pathway. In humans, ER is highly expressed in luminal tumors[1], with the luminal phenotype being characterized by the high expression of some genes that are ER-regulated like *PR* and *RERG[22]*, and other luminal genes that are likely GATA3-regulated including *AGR2* and *K8/18*[46]. In mice, ER expression is low to absent in all the tumors we tested, as is the expression of most human ER-responsive genes. This finding is consistent with previous reports that most late-stage murine mammary tumors are ER-negative ([47] and references within). However, it should be noted that two human luminal tumor-defining genes (*XBP1* and *GATA3*[46], were both highly expressed in murine luminal tumors (Additional data file 2). Taken together, these data suggest that the human "luminal" profile may actually be a combination of at least two profiles, one of which is ER-regulated and another of which is GATA3-regulated; support for a link between *GATA3* and luminal cell origins comes from *GATA3* loss studies in mice where the selective loss of *GATA3* in the mammary gland resulted in either a lack of luminal cells, or a significant decrease in the number and/or maturation of luminal cells[48, 49]. These results suggest that in the mouse models tested here, the ER-regulated gene cassette that is present in human luminal tumors is missing, and that the GATA3-mediated luminal signature remains. Due to the partial luminal tumor signature in mice, we believe that the murine luminal models including TgMMTV-*Neu* profiled here best resemble human luminal tumors and more specifically possibly luminal B tumors, which are luminal tumors that express low amounts of ER and show a poor outcome[2, 3, 21]. While human HER2+/ER- subtype tumors and the murine TgMMTV-*Neu*, TgMMTV-PyMT, and TgWAP-*Myc* fall loosely next to each other in the intrinsic-shared cluster (Figure 2.4), all of the other data argues against this

association. A few murine ER-positive mammary tumor models have been developed[50-53], however, none of these models were analyzed here.

Of note, many expression patterns detected in this study were only observed in one species or the other, and it is possible that some of these differences may arise from technical limitations rather than reflect important biological differences. Comparison between two expression datasets, especially when derived from different species, remains a technical challenge. Thus, we acknowledge the possibility that artifacts may have been introduced depending on the data analysis methodology. However, we are confident that the analyses described here identified many common and biologically relevant clusters including a proliferation, basal epithelial, interferon-regulated and fibroblast signature, thus showing that the act of data combining across species did retain important features present within the individual data sets. These analyses also confirm the notion that there is not a single murine model that perfectly represents a human breast cancer subtype, however, the murine models do show shared features with specific human subtypes and it is these commonalties that will lay the groundwork for many future studies.

MATERIALS AND METHODS

Murine and human tumors.

The murine tumor samples were obtained from multiple participating investigators, who all maintained the mice and harvested the murine tumors in the 0.5-1cm stage following internationally recognized guidelines. The details concerning strain background, promoter, transgene, and specific alleles, *etc*., are provided in Additional data file 1. All human tumor

samples were collected from fresh frozen primary breast tumors using IRB-approved protocols and were profiled as described in[21-23]. The clinical and pathological information for these human samples can be obtained at https://genome.unc.edu/pubsup/breastTumor/.

Microarray Experiments.

Total RNA was collected from murine tumors, and wild type mammary samples of both FVB and BALB/c inbred strains. RNA was purified using the Qiagen RNeasy Mini Kit according to the manufacturer's protocol using 20-30 mg tissue. RNA integrity was assessed using the RNA 6000 Nano LabChip kit followed by analysis using a Bioanalyzer (Agilent). Two micrograms of total RNA was reverse transcribed, amplified and labeled with Cy5 using a Low RNA Input Amplification kit (Agilent). The common reference RNA sample for these experiments consisted of total RNA harvested from equal numbers of C57Bl6/J and 129 male and female Day1 pups (a gift from Dr. Cam Patterson, UNC). The Reference RNA was reverse transcribed, amplified, and labeled with Cy3. The amplified sample and reference were co-hybridized overnight to Agilent Mouse Oligo Microarrays (G4121A). They were then washed and scanned on an Axon GenePix 4000B scanner, analyzed using GenePix 4.1 software and uploaded into our database where a Lowess normalization is automatically performed.

Microarray Data Analysis.

All primary microarray data is available from the University of North Carolina Microarray Database (UMD)[54], and at the Gene Expression Omnibus under the series GSE3165 (mouse and new human data), GSE1992, GSE2740 and GSE2741 (previously

published human data). The genes for all analyses were filtered by requiring the Lowess normalized intensity values in both channels to be > 30. The $\log_2$ ratio of Cy5/Cy3 was then reported for each gene. In the final dataset, only genes that reported values in 70% or more of the samples were included. The genes were median centered and then hierarchical clustering was performed using Cluster v2.12[55]. For the murine unsupervised analysis, and human-mouse unsupervised cluster analyses, we filtered for genes that varied at least 3 fold or more, in at least 3 or more samples. Average linkage clustering was performed on genes and arrays and cluster viewing and display was performed using JavaTreeview v1.0.8[56].

Mouse Intrinsic gene set analysis.

Intrinsic "groups" of experimental samples were chosen based upon having a Pearson correlation value of 0.65 or greater from the unsupervised clustering analysis of the 122 murine samples. The analysis was performed using the Intrinsic Gene Identifier v1.0 by Max Diehn/Stanford University[1]. Technical replicates were removed from the file and the members of every highly correlated node were given identical class numbers, giving every sample that fell outside the 0.65 correlation cut off a class of their own. Using these criteria, 16 groups of samples were identified (see Additional data file 1 for these groups) and a list of 866 "intrinsic" genes was selected using the criteria of one standard deviation below the mean intrinsic gene value. A human intrinsic list of 1300 genes was created using a subset of 146 of the 232 samples used here, and is described in Hu *et al.* 2005[21].

Consensus Clustering.

Consensus Clustering[20] was performed locally using Gene Pattern 1.3.1 (Built Jan 6, 2005), which was downloaded from the Broad Institute distribution site[57]. Analyses

were performed on the mouse dataset with all genes, and just with intrinsic genes separately. Ranges for the number of K clusters (or the focused number of classes) were from 2 to 15 to evaluate a wide range of possible groups. Using a Euclidian distance measure with average linkage, we re-sampled 1000 times with both column and row normalization.

Combining murine and human expression data sets.

Orthologous genes were reported by Mouse Genome Informatics (MGI 3.1) of The Jackson Laboratory. For both the human and murine datasets, Locus Link IDs assigned to Agilent oligo probe ID numbers were used to assign to MGI ID numbers. In cases where a single gene was represented by multiple probes, the median value of the redundant probes was used. This led to a total orthologous pairings of 14,680 Agilent probes. Prior to combining the two datasets, each was column standardized to N(0,1), row median centered, and probe identifiers were converted to MGI IDs. The intersection of mouse and human MGI identifiers from genes that passed filters (same as used above) in both datasets yielded 7907 orthologous genes in the total combined dataset. This data set was next corrected for systemic biases using Distance Weighted Discrimination[24]. Finally, the combined data set was used for an average linkage hierarchical clustering analysis.

Gene Set Enrichment Analysis.

We took the 232 human samples and classified them as Basal-like, Luminal, HER2+/ER-, claudin-low, and Normal Breast-like according to a clustering analysis of the human data set only (not shown), using the new Intrinsic/UNC human gene list developed in Hu *et al*.[21]. Second, the murine samples were also classified based upon their clustering

pattern in Figure 2.1 that used the mouse intrinsic gene list, and were assigned to Groups I-X. Two-class unpaired Significance Analysis of Microarrays analysis was performed for each murine class separately versus all other classes using an FDR of 1%[36], resulting in 10 class-specific gene lists. Using only the set of highly expressed genes that were associated with each analysis (and ignoring the genes whose low expression correlated with a given class), GSEA[35]was performed in R (v. 2.0.1) using the GSEA R package[58]. The 10 murine gene sets were then compared to each human subtype-ranked gene sets and significant enrichments reported. For statistical strength of these enrichments, GSEA uses Family Wise Error Rate (FWER) to correct for multiple testing and False Discovery Rate to reduce false positive reporting. The parameters used for all GSEA were: nperm = 1000, weighted.score.type = 1, nom.p.val.threshold = -1, fwer.p.val.threshold = -1, fdr.q.val.threshold = 0.25, topgs = 12, adjust.FDR.q.val = FALSE, gs.size.threshold.min = 25, gs.size.threshold.max = 2000, reverse.sign = FALSE, preproc.type = 0, random.seed = 3338, perm.type = 0, fraction = 1, replace = FALSE.

Immunofluorescence .

Paraffin-embedded sections (5 μm thick) were processed using standard immunostaining methods. The antibodies and their dilution were α-cytokeratin 5 (K5, 1:8000, Covance, PRB-160P), α-cytokeratins 8/18 (Ker8/18, 1:450 Progen, GP11),. Briefly, slides were deparaffinized and hydrated through a series of xylenes and graded ethanol steps. Heat-mediated epitope retrieval was performed in boiling citrate buffer (pH 6.0) for 15 min., then samples cooled to room temperature for 30 min. Secondary antibodies for immunofluorescence were conjugated with Alexa Fluor-488 or -594 fluorophores (1:200,

Molecular Probes, Invitrogen, Carlsbad, CA). IF samples were mounted with VectaShield Hardset with DAPI mounting media (Vector, Burlingame, CA).

Human KRAS2 amplification assay.

We performed real-time quantitative PCR and fluorescent melting curve analyses using genomic DNAs from 16 Basal-like tumors, a normal breast tissue samples, 2 leukocyte DNA, and 3 luminal tumors. DNA was extracted using the DNAeasy kit (Qiagen) and amplification was performed on the LightCycler using the following temperature parameters: 95ºC, 8m; 50 cycles of 57ºC, 6s; 72ºC, 6s; 95ºC, 2s; followed by cooling to 60ºC and a 0.1ºC/s ramp to 97ºC. Each PCR reaction contained 7.5ng template DNA in a 10µL reaction using the LightCycler Faststart DNA Master SYBR Green I kit (Roche). Relative DNA copy number for each gene was determined by importing an external efficiency curve and using a "normal" breast sample for a within-run calibrator. For each sample, the copy number for KRAS2 was divided by the average copy number of ACTB and G1P3. Amplification in any tumor was called if the relative fold change was greater than 3 standard deviations above the average of 5 control samples (2 normal leukocyte samples and 3 luminal tumors).

ACKNOWLEDGEMENTS

C.M.P. was supported by funds from the NCI Breast SPORE program to UNC-CH (P50-CA58223-09A1), by NCI (RO1-CA-101227-01), by the Breast Cancer Research Foundation and by HHSN-261200433008C (N01-CN43308). K.S. was supported by a grant from NIEHS (T32 ES07017) and T.V.D was supported NCI (RO1-CA046283-16). R.G. was supported by NO1-CN15044, and P.A.F. was supported by NO1-CN-05024/CN/NCI. This

research was supported in part by a grant from the Susan G. Komen Breast Cancer Foundation (LPJ and SA). P.H.B. was supported by funds from RO1-CA101211. We thank Beverly H. Koller and Daniel Medina for generously providing tumor samples, and Ronald Lubet for assistance in obtaining murine tumor samples.

REFERENCES

1. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**(6797):747-752.

2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc Natl Acad Sci U S A* 2001, **98**(19):10869-10874.

3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: **Repeated observation of breast tumor subtypes in independent gene expression data sets**. *Proc Natl Acad Sci U S A* 2003, **100**(14):8418-8423.

4. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P *et al*: **Breast cancer molecular subtypes respond differently to preoperative chemotherapy**. *Clin Cancer Res* 2005, **11**(16):5678-5685.

5. Troester MA, Hoadley KA, Sorlie T, Herbert BS, Borresen-Dale AL, Lonning PE, Shay JW, Kaufmann WK, Perou CM: **Cell-type-specific responses to chemotherapeutics in breast cancer**. *Cancer Res* 2004, **64**(12):4218-4226.

6. Van Dyke T, Jacks T: **Cancer modeling in the modern era: progress and challenges**. *Cell* 2002, **108**(2):135-144.

7. Hennighausen L: **Mouse models for breast cancer**. *Oncogene* 2000, **19**(8):966-967.

8. Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J, Matusik R, Thomas GV, Sawyers CL: **Myc-driven murine prostate cancer shares molecular features with human prostate tumors**. *Cancer Cell* 2003, **4**(3):223-238.

9. Lee JS, Chu IS, Mikaelyan A, Calvisi DF, Heo J, Reddy JK, Thorgeirsson SS: **Application of comparative functional genomics to identify best-fit mouse models to study human cancer**. *Nat Genet* 2004, **36**(12):1306-1311.

10. Sweet-Cordero A, Mukherjee S, Subramanian A, You H, Roix JJ, Ladd-Acosta C, Mesirov J, Golub TR, Jacks T: **An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis**. *Nat Genet* 2005, **37**(1):48-55.

11. Cano A, Perez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, Portillo F, Nieto MA: **The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression**. *Nat Cell Biol* 2000, **2**(2):76-83.

12. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR *et al*: **Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome**. *Am J Pathol* 2002, **161**(6):1991-1996.

13. Colnot DR, Nieuwenhuis EJ, Kuik DJ, Leemans CR, Dijkstra J, Snow GB, van Dongen GA, Brakenhoff RH: **Clinical significance of micrometastatic cells detected by E48 (Ly-6D) reverse transcription-polymerase chain reaction in bone marrow of head and neck cancer patients**. *Clin Cancer Res* 2004, **10**(23):7827-7833.

14. Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, Ferno M, Peterson C, Meltzer PS: **Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns**. *Cancer Res* 2001, **61**(16):5979-5984.

15. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET: **Breast cancer classification and prognosis based on gene expression profiles from a population-based study**. *Proc Natl Acad Sci U S A* 2003, **100**(18):10393-10398.

16. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT *et al*: **Gene expression profiling predicts clinical outcome of breast cancer**. *Nature* 2002, **415**(6871):530-536.

17. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA, Jr., Marks JR, Nevins JR: **Predicting the clinical status of human breast cancer by using gene expression profiles**. *Proc Natl Acad Sci U S A* 2001, **98**(20):11462-11467.

18. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, Perou CM: **Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma**. *Mod Pathol* 2005.

19. Li Y, Hively WP, Varmus HE: **Use of MMTV-Wnt-1 transgenic mice for studying the genetic basis of breast cancer**. *Oncogene* 2000, **19**(8):1002-1009.

20. Monti S, Tamayo P, Mesirov J, Golub TR: **Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data**. *Machine Learning* 2003, **52**(1-2):91-118.

21. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L *et al*: **The molecular portraits of breast tumors are conserved across microarray platforms**. *BMC Genomics* 2006, **7**(1):96.

22. Oh DS, Troester MA, Usary J, Hu Z, He X, Fan C, Wu J, Carey LA, Perou CM: **Estrogen-Regulated Genes Predict Survival in Hormone Receptor-Positive Breast Cancers**. *J Clin Oncol* 2006.

23. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, Glas AM, Perou CM, Van't Veer LJ: **Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer**. *Cancer Res* 2005, **65**(20):9155-9158.

24. Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, Marron JS: **Adjustment of systematic microarray data biases**. *Bioinformatics* 2004, **20**(1):105-114.

25. Moyano JV, Evans JR, Chen F, Lu M, Werner ME, Yehiely F, Diaz LK, Turbin D, Karaca G, Wiley E *et al*: **alphaB-Crystallin is a novel oncoprotein that predicts poor clinical outcome in breast cancer**. *J Clin Invest* 2006, **116**(1):261-270.

26. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L *et al*: **Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma**. *Clin Cancer Res* 2004, **10**(16):5367-5374.

27. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC *et al*: **Distinctive gene expression patterns in human mammary epithelial cells and breast cancers**. *Proc Natl Acad Sci U S A* 1999, **96**(16):9212-9217.

28. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO *et al*: **Identification of genes periodically expressed in the human cell cycle and their expression in tumors**. *Mol Biol Cell* 2002, **13**(6):1977-2000.

29. Liu ML, Von Lintig FC, Liyanage M, Shibata MA, Jorcyk CL, Ried T, Boss GR, Green JE: **Amplification of Ki-ras and elevation of MAP kinase activity during mammary tumor progression in C3(1)/SV40 Tag transgenic mice**. *Oncogene* 1998, **17**(18):2403-2411.

30. Liu ML, Shibata MA, Von Lintig FC, Wang W, Cassenaer S, Boss GR, Green JE: **Haploid loss of Ki-ras delays mammary tumor progression in C3 (1)/SV40 Tag transgenic mice**. *Oncogene* 2001, **20**(16):2044-2049.

31. van Beers EH, van Welsem T, Wessels LF, Li Y, Oldenburg RA, Devilee P, Cornelisse CJ, Verhoef S, Hogervorst FB, van't Veer LJ *et al*: **Comparative genomic hybridization profiles in human BRCA1 and BRCA2 breast tumors highlight differential sets of genomic aberrations**. *Cancer Res* 2005, **65**(3):822-827.

32. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, Trudel M, Akslen LA: **Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer**. *J Natl Cancer Inst* 2003, **95**(19):1482-1485.

33. Ross DT, Perou CM: **A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines**. *Dis Markers* 2001, **17**(2):99-109.

34. Elenbaas B, Spirio L, Koerner F, Fleming MD, Zimonjic DB, Donaher JL, Popescu NC, Hahn WC, Weinberg RA: **Human breast cancer cells generated by oncogenic**

transformation of primary mammary epithelial cells. *Genes Dev* 2001, **15**(1):50-65.

35. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci U S A* 2005, **102**(43):15545-15550.

36. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response**. *Proc Natl Acad Sci U S A* 2001, **98**(9):5116-5121.

37. Simin K, Wu H, Lu L, Pinkel D, Albertson D, Cardiff RD, Dyke TV: **pRb Inactivation in Mammary Cells Reveals Common Mechanisms for Tumor Initiation and Progression in Divergent Epithelia**. *PLoS Biol* 2004, **2**(2):E22.

38. Hennighausen LG, Sippel AE: **Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins**. *Nucleic Acids Res* 1982, **10**(8):2677-2684.

39. Munoz B, Bolander FF, Jr.: **Prolactin regulation of mouse mammary tumor virus (MMTV) expression in normal mouse mammary epithelium**. *Mol Cell Endocrinol* 1989, **62**(1):23-29.

40. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, Chinnaiyan AM: **Mining for regulatory programs in the cancer transcriptome**. *Nat Genet* 2005, **37**(6):579-583.

41. Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, Mone M, Hansen H, Buys SS, Rasmussen K *et al*: **Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay**. *Breast Cancer Res* 2006, **8**(2):R23.

42. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J *et al*: **Gene Expression and Benefit of Chemotherapy in Women With Node-Negative, Estrogen Receptor-Positive Breast Cancer**. *J Clin Oncol* 2006.

43. Crook T, Brooks LA, Crossland S, Osin P, Barker KT, Waller J, Philp E, Smith PD, Yulug I, Peto J *et al*: **p53 mutation with frequent novel condons but not a mutator phenotype in BRCA1- and BRCA2-associated breast tumours**. *Oncogene* 1998, **17**(13):1681-1689.

44. Phillips KA, Nichol K, Ozcelik H, Knight J, Done SJ, Goodwin PJ, Andrulis IL: **Frequency of p53 mutations in breast carcinomas from Ashkenazi Jewish carriers of BRCA1 mutations**. *J Natl Cancer Inst* 1999, **91**(5):469-473.

45. Li Y, Welm B, Podsypanina K, Huang S, Chamorro M, Zhang X, Rowlands T, Egeblad M, Cowin P, Werb Z *et al*: **Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells**. *Proc Natl Acad Sci U S A* 2003, **100**(26):15853-15858.

46. Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, Langerod A, Karesen R, Oh DS, Dressler LG *et al*: **Mutation of GATA3 in human breast tumors**. *Oncogene* 2004, **23**(46):7669-7678.

47. Medina D: **Mammary developmental fate and breast cancer risk**. *Endocr Relat Cancer* 2005, **12**(3):483-495.

48. Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z: **GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland**. *Cell* 2006, **127**(5):1041-1055.

49. Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, Hartley L, Robb L, Grosveld FG, van der Wees J *et al*: **Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation**. *Nat Cell Biol* 2006.

50. Wijnhoven SW, Zwart E, Speksnijder EN, Beems RB, Olive KP, Tuveson DA, Jonkers J, Schaap MM, van den Berg J, Jacks T *et al*: **Mice expressing a mammary gland-specific R270H mutation in the p53 tumor suppressor gene mimic human breast cancer development**. *Cancer Res* 2005, **65**(18):8166-8173.

51. Frech MS, Halama ED, Tilli MT, Singh B, Gunther EJ, Chodosh LA, Flaws JA, Furth PA: **Deregulated estrogen receptor alpha expression in mammary epithelial cells**

of transgenic mice results in the development of ductal carcinoma in situ. *Cancer Res* 2005, **65**(3):681-685.

52. Lin SC, Lee KF, Nikitin AY, Hilsenbeck SG, Cardiff RD, Li A, Kang KW, Frank SA, Lee WH, Lee EY: **Somatic mutation of p53 leads to estrogen receptor alpha-positive and -negative mouse mammary tumors with high frequency of metastasis**. *Cancer Res* 2004, **64**(10):3525-3532.

53. Torres-Arzayus MI, De Mora JF, Yuan J, Vazquez F, Bronson R, Rue M, Sellers WR, Brown M: **High tumor incidence and activation of the PI3K/AKT pathway in transgenic mice define AIB1 as an oncogene**. *Cancer Cell* 2004, **6**(3):263-274.

54. **The UNC Microarray Database** [www.genome.unc.edu]

55. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns**. *Proc Natl Acad Sci U S A* 1998, **95**(25):14863-14868.

56. Saldanha AJ: **Java Treeview--extensible visualization of microarray data**. *Bioinformatics* 2004, **20**(17):3246-3248.

57. **Gene Pattern 1.3** [http://www.broad.mit.edu/cancer/software/genepattern/]

58. **Gene Set Enrichment Analysis (GSEA)** [http://www.broad.mit.edu/personal/aravind/GSEA/]

CHAPTER III

FURTHER MOLECULAR PROFILING OF MOUSE MAMMARY TUMOR MODELS

ABSTRACT

We previously compared the expression profiles among 13 mouse mammary tumor models and compared them to the human intrinsic molecular subtypes of breast cancer. This study has been continued and we now present data corresponding to more than 20 mouse models. In our original study we did not include any ER+ models because unlike human breast cancers and rat carcinogen-induced mammary tumors, ER+ mammary tumors are uncommon in the mouse. We have now profiled several ER+ models and in our analysis, these models do not mimic human ER+ luminal subtype tumors and do not highly express ER target genes. Many of these models belong to a newly identified mouse molecular subtype with alveolar expression features. A previously derived gene expression signature of p53 function shows high expression in mouse models with perturbed p53 further validating this gene set and the use of these models. Finally, we find that tumors that arise in the mouse mammary gland due to haploinsufficiency for Brg1 span across most mouse tumor types indicating that Brg1 may have a possible role in the mammary gland stem or progenitor cell function.

INTRODUCTION

Genetic engineering of mice to study mammary biology is now in its third decade. Well over 100 transgenes, targeted mutations, and compound models (combinations of transgenes and combinations of transgenes and targeted mutations) with increasing complexity have been used to study mammary tumorigenesis in mice. We recently reported the comparison between 13 mouse mammary tumor models and 200 human breast tumors using gene expression profiling[1]. We showed that although no single mouse model recapitulated all the expression features of a given human subtype, many of the defining characteristics of the human subtypes were conserved among the mouse models. Here we report a continuation of these profiling efforts using 10 additional murine tumor models.

ER-alpha (ESR1) expression is one of the most common features of human breast cancers with ~70% of tumors reported to be positive. Strikingly, mouse mammary models are almost entirely ER-negative in contrast to chemically induced rats which show nearly 100% positivity. In our original study, mouse models known to be ER+ were not included, however, now we have profiled several ER+ models: these include TgMMTV-*AIB1* in which amplified in breast cancer 1, an estrogen receptor coactivator, is overexpressed under the MMTV promoter[2]. TgWap-Cre; Tel-NTRK3 mice conditionally express the Tel(Etv6)-NTRK3 fusion oncoprotein, a translocation characteristic of secretory breast cancer (SBC), a rare form of breast cancer. TgWap-Cre; p53flox/flox mice develop predominantly ER-positive mammary tumors[3]. p18-/- mice on the BALBc genetic background develop mammary tumors with a relatively long latency and have been described as ER+. When this

model is crossed to BRCA1+/- mice to generate p18-/-; BRCA1+/- mice, the latency decreases and the tumors have been reported to be predominantly ER-negative. In addition, some TgMMTV-*Wnt1* tumors have been reported to be ER-alpha positive[4]. Surprisingly, in our analysis, all of these ER+ murine models do not seem to closely mimic human ER+ luminal tumors. Many of these models instead belong to a newly identified mouse molecular subtype with alveolar expression features and this murine subtype does not appear to have a clear human counterpart. A previously derived gene expression signature of p53 function shows high expression in mouse models with perturbed p53 further validating this gene set and the use of these models[5]. We also find that tumors that arise in the mouse mammary gland due to haploinsufficiency for Brg1 span across most mouse tumor expression subtypes indicating that Brg1 may have a possible role in the mammary gland stem or progenitor cell[6].

Table 3.1. Summary of mouse mammary tumor models.

Tumor Model	Tumor number	Promoter	Transgene/Gene knockout	Strain	Reference
WAP-myc	20	WAP	Myc overexpression	FVB	[7]
WAP-Int3	7	WAP	Notch4 overexpression	FVB	[8]
WAP-T_{121}	5	WAP	pRb,p107,p130 inactivation	B6D2	[9]
WAP-T_{121}	2	WAP	pRb,p107,p130 inactivation	BALB/cJ	[9]
WAP-Tag	5	WAP	SV40 L-T (pRb, p107, p130, p53, p300 inactivation, others); SV40 s-t	C57Bl/6	[10]
C3(1)-Tag	16	C3(1)	SV40 L-T (pRb, p107, p130, p53, p300 inactivation, others); SV40 s-t	FVB	[11]
MMTV-neu	18	MMTV	Unactivated rat Her2 overexpression	FVB	[12]
MMTV-Wnt1	15	MMTV	Wnt 1 overexpression	FVB	[13]
MMTV-PyMT	7	MMTV	PyMT (activation of Src, PI-3' kinase, and Shc)	FVB	[14]
MMTV-Cre;$Brca1^{Co/Co}$; $p53^{+/-}$	10	MMTV	Brca1 truncation mutant; p53 heterozygous null	C57Bl/6	[15]
$p53^{-/-}$ transplanted	5	none	p53 inactivation	BALB/cJ	[16]
Medroxyprogesterone-DMBA-induced	11	none	Random DMBA-induced	FVB	[17]
$p53^{+/-}$ irradiated	7	none	p53 heterozygous null, random IR induced	BALB/cJ	[18]
$Brca1^{+/-}$;$p53^{+/-}$ irradiated	7	none	Brca1 and p53 heterozygous null, random IR induced	BALB/cJ	[19]
MMTV-Fgf3	5	MMTV	Fgf3/int2 overexpression	FVB	[20, 21]
MMTV-DN89β-catenin	4	MMTV	Overexpression of activated β-catenin	FVB	[22]
WAP-Cre; p53flox/flox	4	WAP	p53 inactivation	C57Bl/6 129	[3]
$p53^{-/-}$ spontaneous	5	none	p53 inactivation	BALB/cJ	[23]

p18$^{-/-}$	5	none	p18 inactivation	BALB/cJ	unpublished
p18$^{-/-}$; Brca1$^{+/-}$	12	none	p18 inactivation	BALB/cJ	unpublished
MMTV-AIB1	9	MMTV	AIB1 overexpression	FVB	[2]
WAP-Cre; Tel-NTRK3	12	WAP	Tel-NTRK3 fusion protein expressed from endogenous Tel 1 locus	mixed	unpublished
Brg1$^{+/-}$	12	none	Brg1heterozygous null	mixed	[6]
MMTV-HRAS	8	MMTV	Overexpression of activated HRAS	FVB	[24]

Tumor models in bold are new to this study.

RESULTS

Previous intrinsic groups are retained

We included the previously published 122 microarrays consisting of 13 tumor models and normal mammary glands from BALB/c and FVB/N mice and added an additional 113 microarrays. This additional set included 10 new models and 3 normal mammary glands from C57BL6 mice (Table 3.1). We performed a new "intrinsic" analysis to select genes consistently representative of groups/classes of murine samples. As in the previous analysis, we applied the algorithm to groups of murine samples defined by an empirically determined correlation threshold of > 0.65 using the unsupervised cluster dendrogram. This "intrinsic" analysis yielded 1802 genes that we then used in a hierarchical cluster analysis (Figure 3.1). This analysis again identified the same nine tumor groups (II-X) from the first analysis as well as the normal mammary gland group (Group I). There were an additional two large tumor groups which we designated groups XI and XII. Group I contained all the normal mammary gland samples including the newly added C57BL6 samples. This cluster also now

included some tumors in the outer branches (1/5 *p18*$^{-/-}$, 2/5 TgMMTV-*Fgf3*, 1/18 TgMMTV-*Neu*). Normal tissue contamination is most likely responsible for the clustering of tumor samples in this group.

Group II samples generally have a spindloid morphology and show high expression of mesenchymal genes (previously shown). These were derived from several models (2/10 *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/-}$, 3/11 DMBA-induced, 1/5 *p53*$^{-/-}$ transplant, 1/7 *p53*$^{+/-}$ IR, 1/18 TgMMTV-*Neu* and 1/7 TgWAP-T_{121}). New to this cluster was a single *p53*$^{-/-}$ spontaneous tumor (1/5), a TgMMTV- *AIB1* tumor (1/9), and a tumor from a Brg1$^{+/-}$ mouse (1/12).

Group III contains 2/11 DMBA-induced and 9/16 TgMMTV- *Wnt1*. Interestingly, Group III now also contains all 4 TgMMTV-DN89β-catenin tumors and a TgMMTV-*Fgf3* tumor (1/5). Two other TgMMTV-*Fgf3* tumors are clustered just outside this group. β-catenin is a key downstream mediator of the wnt signaling pathway and mammary tumors from β-catenin transgenic mice have been shown to histologically resemble TgMMTV-*Wnt1* tumors[25, 26]. Fgf3, interestingly, has been shown to cooperate with Wnt1 during mammary tumorigenesis[20, 27].

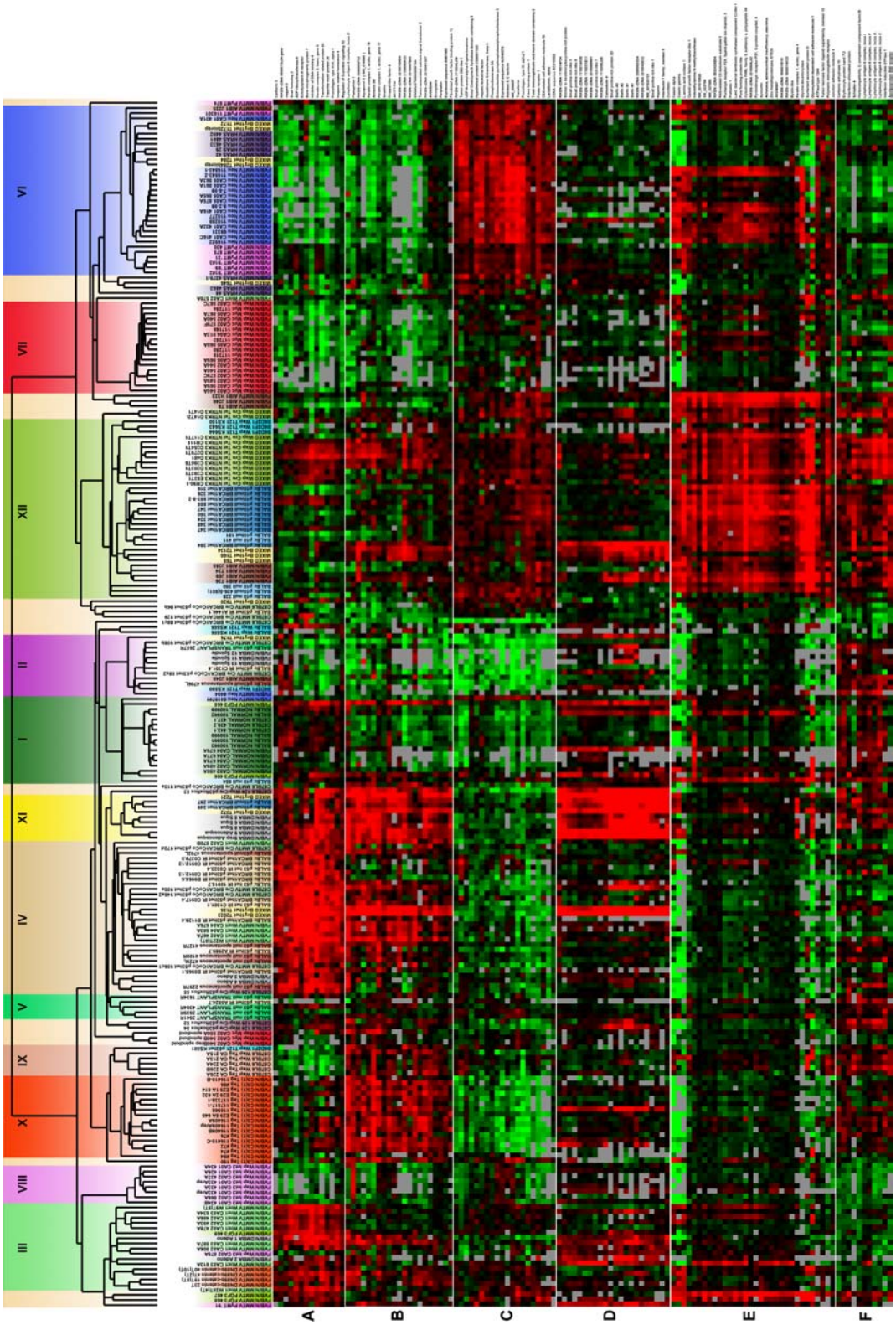

VI
VII
XII
II
I
XI
IV
V
IX
X
VIII
III
A
B
C
D
E
F

Figure 3.1 Intrinsic clustering of mouse mammary tumor models. The dendrogram is colored to indicate 12 groups. (A) Cluster of basal genes including *Keratin 5*. (B) A second basal epithelial cluster of genes including *Keratin 14* and *17*. (C) XBP1 luminal epithelial gene expression cluster. (D) squamous gene cluster. (E) alveolar gene expression cluster. (F) A cluster containing Sca-1/Ly-6A and other Ly-6 family members.

Group IV contains 7/7 Brca1$^{+/-}$;*p53*$^{+/-}$ IR, 4/10 *Brca1*$^{Co/Co}$;TgMMTV-Cre;*p53*$^{+/-}$, 4/7 *p53*$^{+/-}$ IR, 2/11 DMBA-induced, and 5/16 TgMMTV- *Wnt1*. New to this data set, this cluster now includes 1/4 TgWap-Cre; p53flox/flox tumor, 2/12 Brg1$^{+/-}$ tumors, and all 5 p53$^{-/-}$ spontaneous tumors. As we previously reported, Groups III and IV are characterized by the expression of a gene cluster that includes Keratin 5, P-cadherin, and ID4 (Figure 3.1A). They also show high expression of a Keratin 14, Keratin 17, and Ly-6D cluster although not as uniformly (Figure 3.1B). Group V, as before, contains 4/5 *p53*$^{-/-}$ transplant tumors and a single *p53*$^{+/-}$ IR tumor (1/7). Why these tumors, arising from *p53*$^{-/-}$ preneoplastic epithelium transplanted into the cleared mammary fat pad of syngeneic animals, cluster separately from the *p53*$^{-/-}$ tumors arising spontaneously is not known and likely merits further investigation.

Table 3.2 Group III and IV Wnt1 tumors exhibit a difference in tumor latency.

Tumor	Transgene	DOB	Date of Necropsy	Age (mo)	Array tumor type	Breeding Status
CA02-467A	Wnt1	1/23/2002	3/21/2002	1.9	Group IV	Virgin
CA02-478A	Wnt1	12/9/2001	5/9/2002	5.0	Group III	Virgin
CA02-486A	Wnt1	1/25/2002	6/13/2002	4.6	Group III	Virgin
CA02-493A	Wnt1	1/22/2002	6/20/2002	5.0	Group III	Virgin
CA02-506A	Wnt1	10/15/2001	4/18/2002	6.2	Group III	Virgin
CA02-570A	Wnt1	N/A	12/6/2002	N/A	no group (luminal)	Virgin
CA02-570B	Wnt1	N/A	12/6/2002	N/A	Group IV	Virgin
CA03-587A	Wnt1	9/10/2002	2/5/2003	4.9	Group III	Virgin
CA03-613A	Wnt1	N/A	4/23/2003	N/A	Group III	Virgin
CA03-634A	Wnt1	8/25/2003	1/6/2004	4.5	Group III	Virgin
CA04-676A	Wnt1	2/9/2004	4/22/2004	2.4	Group IV	Virgin
CA04-683A	Wnt1	3/2/2004	6/3/2004	3.1	Group IV	N/A

Interestingly, as reported previously, tumors from the TgMMTV- *Wnt1* model were split almost exclusively into these two groups (Groups III and IV). Examining the latency data we have available for a subset of these tumors coming from one institution, it was

apparent that there was a latency difference between Wnt1 tumors in the two groups with the Group III Wnt1 tumors (N=6) showing latencies of 4.5-6.2 months and the Group IV tumors (N=3) between 1.9-3.1 months (Table 3.2). Furthermore, tumors in the two groups appear to have histological differences with tumors in Group IV showing a distinct radiating architecture (Figure 3.2). One key expression difference between the two classes of Wnt1 tumors is that the Group III Wnt1 tumors have a low expression of the XBP1 luminal gene cluster (Figure 3.1C). Tumors that arise from crosses between TgMMTV- *Wnt1* and TgMMTV-*Neu* or $p53^{+/-}$ cluster almost exclusively in Group IV (data not shown). Further studies using the Wnt1 mice now in our colony are warranted to validate these results.

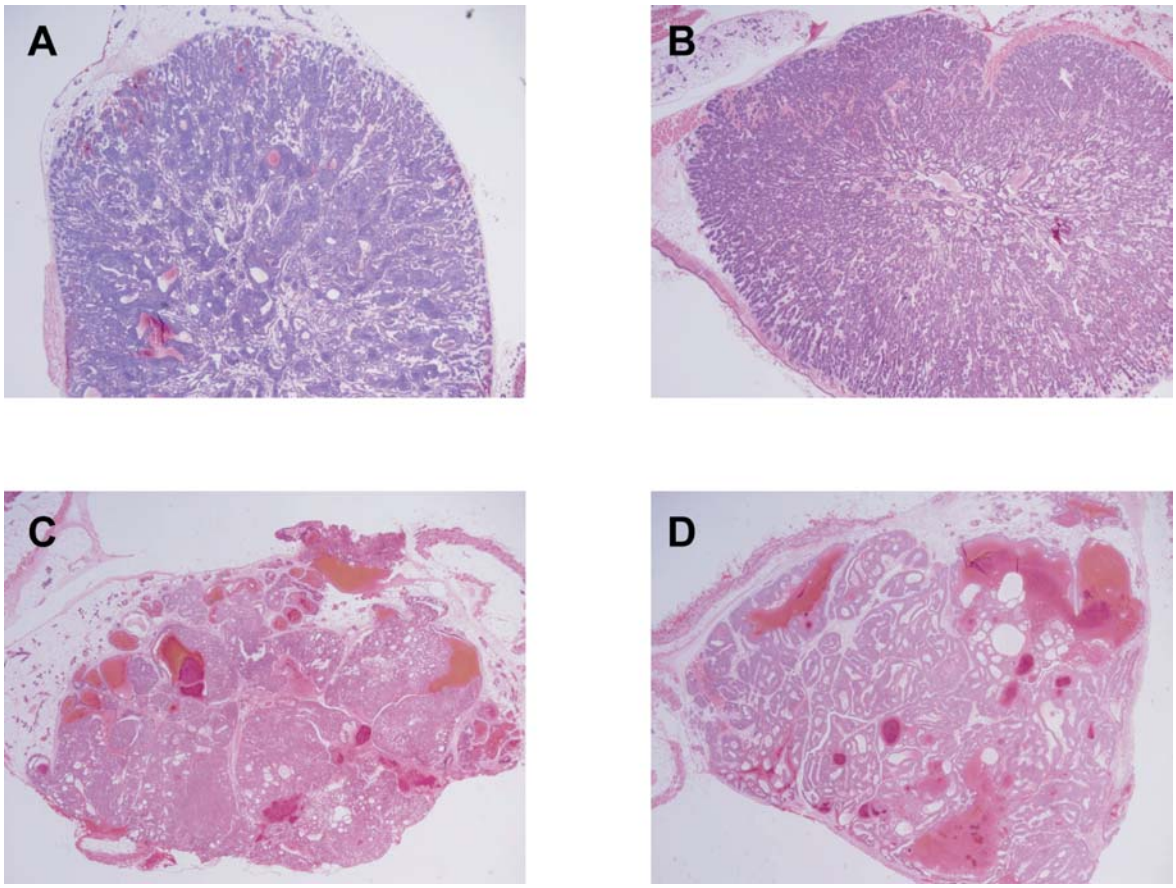


Figure 3.2 Group III and IV Wnt1 tumors show distinct histologies. Group IV Wnt1 tumors (A,B) show a distinct radiating pattern with terminal end buds in the periphery and

ducts in the center different from Group III Wnt1 tumors (C,D). Pictures courtesy of Igor Mikaelian.

Group VI contains 16/18 TgMMTV-*Neu* tumors and 7/9 TgMMTV-*PyMT* tumors. In addition, this cluster now contains 2/12 $Brg1^{+/-}$ tumors and 5/8 mammary tumors arising in TgMMTV-HRAS mice. TgMMTV-HRAS tumors have been shown previously to be similar to TgMMTV-*Neu* and TgMMTV-*PyMT* both histologically and by gene expression[26, 28]. These tumors are characterized by the high expression of the previously reported luminal XBP1 cluster (Figure 3.1C). Groups VII and VIII again contained most of the TgWAP-*Myc* tumors (18/20) and TgWAP-*Int3* samples (6/7), respectively. Group IX contains 1/7 TgWAP-T_{121} tumors and all TgWAP-*Tag* tumors and Group X contains all 16 TgC3(1)-*Tag* tumors. These two groups show the highest expression of proliferation genes (previously reported) and high expression of the Keratin 14 gene cluster (Figure 3.1B).

Our previous analysis included a few samples described as showing extensive squamous differentiation. In the previous report, however, we did not refer to this as a group because of the small number of samples contained. This group now contains 8 tumors, which we labeled Group XI. These samples include 4/11 DMBA-induced tumors, 2/12 $Brg1^{+/-}$ tumors, and 2/12 $p18^{-/-}$; $BRCA1^{+/-}$ tumors. In addition to the high expression of the Keratin 5 and Keratin 14 clusters (Figure 3.1A,B), these tumors show very high expression of a cluster of genes that includes many genes involved in squamous differentiation (Figure 3.1D). Cornifelin is part of the insoluble cornified cell envelope of stratified squamous epithelia and Small proline-rich proteins are precursor proteins for the crosslinked envelope

formation in cells undergoing squamous differentiation. The expanded cluster also contains many other keratin genes.

A new subtype of mouse mammary tumor

Group XII is a new cluster containing a large number of the newly included samples/models, many of them reported to be ER-positive. This cluster contains 4/5 $p18^{-/-}$ tumors, 10/12 $p18^{-/-}$ $BRCA1^{+/-}$ tumors, 4/9 TgMMTV-*AIB1* tumors, 10/12 TgWAP-Cre; Tel-NTRK3 tumors, 3/7 TgWAP-T121 and 3/12 $Brg1^{+/-}$ tumors. Interestingly, common to this group was expression of a large cluster of genes including milk proteins, casein-alpha and gamma (Figure 3.1E), whey acidic protein, and MUC1 (not shown). These tumors also highly express the murine hematopoetic stem cell antigen, Sca-1/Ly-6A and other Ly-6 family members (Figure 3.1F). These tumors also express the luminal XBP1 cluster at lower levels than Group VI. This group is also heterogeneous with most of the TgWAP-Cre; Tel-NTRK3 tumors and some of the $p18^{-/-}$ $BRCA1^{+/-}$ tumors expressing the Keratin 5 gene cluster (Figure 3.1A). A number of the tumors in this group including most of the TgWAP-Cre; Tel-NTRK3 also express the Keratin 14 gene cluster (Figure 3.1B). Unlike Group VI, Group XII tumors did not show high expression of GATA-3, however, they highly express another GATA family member, GATA-2 (not shown). It is interesting to speculate that GATA-2 may have a similar role in alveolar cell fate and differentiation as GATA-3 has been shown to have in ductal luminal cells.

Brg1+/- mice give rise to tumors that span across most tumor types

Brg1 tumors were found throughout the intrinsic cluster with at least one of these tumors being found in 5 of the 11 tumor groups including the new group with alveolar features (i.e. Group XII). We speculate that Brg1 haploinsufficiency perturbs a mammary gland stem or multipotent progenitor in such a way that the secondary genetic changes that occur then committed the cells into the various subtypes.

p53 signature

We previously developed a p53 gene signature by comparing human cell line signatures of p53 loss with p53-mutation associated genes from human breast tumors[5]. Here, we used the murine orthologs of these genes and averaged their expression to investigate the importance of the human p53 signature in our murine models (Figure 3.3). Strikingly, the "p53 up" signature appears to be higher in most murine tumors with deficient p53 function including both T-antigen models and at the same time the "p53 down" signature is low. Interestingly, many of the TgWAP-*Myc* tumors also show this signature. No studies looking for p53 mutations in this model have been published. Whether this observation is a function of p53 pathway inactivation or cross-talk with Myc remains to be determined.

Mouse mammary tumors do not highly express the ER response signature

Similar to our approach with the p53 signature, we used a previously developed estrogen regulated gene signature that was developed by comparing the response of MCF-7 to 17-beta estradiol filtered through ER+ human breast tumors[29]. We used the murine

orthologs of these genes and again averaged their expression. Not many, if any, of the mouse tumors appear to highly express these genes. With the exception of a single Wnt1 tumor (613A), the normal mammary glands seemed to have higher expression than the tumors. This seems to indicate that murine tumors do not express the ER-regulated genes we see in human tumors. Further methods need to be used to validate this finding.

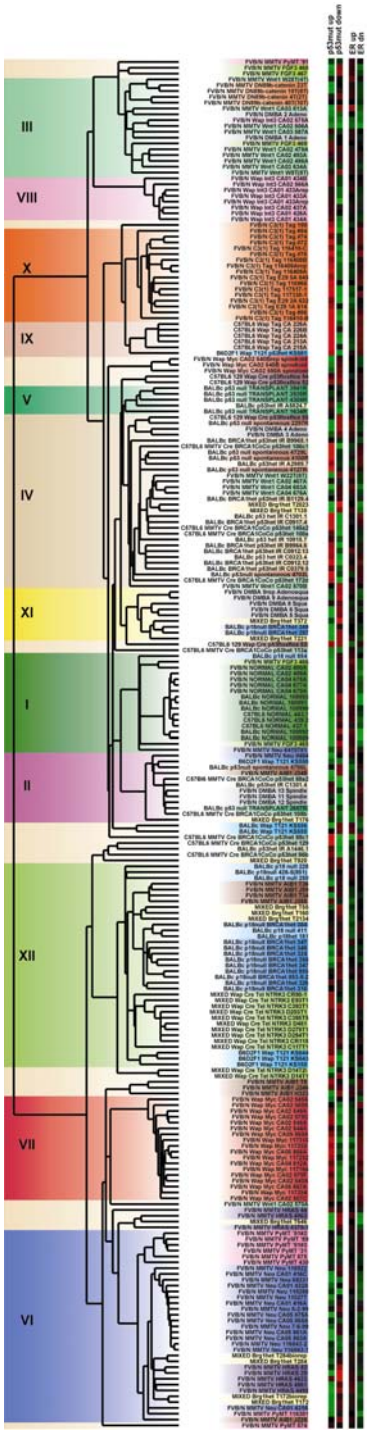


Figure 3.3 Pathway analysis in mouse models of breast cancer. The samples are clustered by the intrinsic gene list as in Figure 3.1. The previously developed p53 and ER response genes were separated into up and down genes were standardized, median centered, and the profile was averaged for each experimental sample.

Mouse-human comparison

As we had done in the previous report, we compared gene lists derived from the mouse classes with the human subtypes using Gene Set Enrichment Analysis (GSEA) with similar results (Table 3.3). We again found that the murine Groups IX (p=0.006) and X (p=0.001), which were comprised of tumors from T-antigen initiated models, shared significant overlap with the human basal-like subtype. The murine Group VI (TgMMTV-*Neu*, TgMMTV-*PyMT*, and TgMMTV-*HRAS*) showed a now significant association (p=0.04) with the human luminal profile. The murine Group II spindloid tumors showed significant overlap with claudin-low tumors (p=0.001). Group XII (alveolar group) did not show significant gene enrichment with any of the human tumor subtypes. It was the closest with the HER2+/ER- subtype (NOM p=0.02, FWER p=0.25).

Table 3.3 Gene Set Enrichment Analysis (GSEA) of the 11 murine groups versus 5 human subtypes. Significant values are in bold. Group V was not analyzed because a SAM FDR of 1% resulted in no significant genes.

		Basal		Luminal		HER2		Claudin-low		Normal	
Group	Size	NOM p-value	FWER p-value	NOM p-value	FWER p-value	NOM p-value	FWER p-value	NOM p-value	FWER p-value	NOM p-value	FWER p-value
I	1313	-	-	0.213	0.654	0.638	0.969	0.018	0.111	0.251	0.666
II	724	-	-	-	-	0.714	0.996	0.002	**0.001**	-	-
III	457	-	-	-	-	0.433	0.893	-	-	0.002	0.059
IV	790	0.015	0.218	-	-	-	-	-	-	0.322	0.815

VI	1411	-	-	0.002	**0.04**	0.51	0.919	-	-	-	-
VII	1036	0.029	0.103	-	-	-	-	-	-	-	-
VIII	112	0.072	0.429	-	-	-	-	0.156	0.586	0.066	0.359
IX	156	0.006	**0.006**	-	-	-	-	-	-	-	-
X	665	0	**0.001**	-	-	-	-	0.75	0.996	-	-
XI	209	0.773	0.997	-	-	0.289	0.72	0.015	0.059	-	-
XII	527	-	-	0.579	0.94	0.016	0.258	-	-	-	-

		NOT Basal		NOT Luminal		NOT HER2		NOT Claudin-low		NOT Normal	
Group	Size	NOM p-value	FWER p-value	NOM p-value	FWER p-value	NOM p-value	FWER p-value	NOM p-value	FWER p-value	NOM p-value	FWER p-value
I	1313	0.006	**0.029**	-	-	-	-	-	-	-	-
II	724	0.426	0.895	0.915	0.999	-	-	-	-	0.435	0.914
III	457	0.661	0.975	0.654	0.973	-	-	0.859	0.999	-	-
IV	790	-	-	0.282	0.785	0.419	0.883	0.372	0.877	-	-
VI	1411	0	**0.043**	-	-	-	-	0.067	0.371	0.998	0.215
VII	1036	-	-	0.13	0.438	0.621	0.977	0.286	0.787	0.37	0.835
VIII	112	-	-	0.203	0.659	0.11	0.48	-	-	-	-
IX	156	-	-	0.106	0.161	0.889	1	-	-	0.111	0.148
X	665	-	-	0.01	**0.047**	0.65	0.981	-	-	0.096	0.307
XI	209	-	-	0.438	0.887	-	-	-	-	0.225	-
XII	527	0.508	0.938	-	-	-	-	0.044	0.303	0.314	0.83

DISCUSSION

ER-alpha expression is one of the most common features of human breast cancers with ~70% of tumors reported to be positive. Mouse models, on the other hand, are almost entirely ER-negative. In our original study, mouse models known to be ER+ were not included. We have now profiled several ER+ models. None of these models showed high expression of a previously published ER-response signature, however, additional analyses are needed to confirm the significance of this negative finding. Many of these tumors are, however, part of a new mouse mammary tumor subtype with an alveolar phenotype.

The p53 tumor suppressor is the most commonly altered gene in human breast cancer. It has been found to be mutated in about 30-40% of all human breast tumors with much higher frequency in the two ER negative subtypes[30]. p53 plays an important role in protecting the genome from DNA damage via blocking the cell cycle in the G1 phase, such that the cell is able to repair genomic damage. If the cell is beyond repair, p53 can induce apoptosis or senescence. When p53 function is removed, aberrant cell growth continues unchecked and the cells become genomically unstable. In a large number of the mouse models we have included in our study, p53 is perturbed, sometimes it is the initiating event ($p53^{-/-}$ transplant or spontaneous and $p53^{+/-}$; IR) and others it is in combination with other initiators ($Brca1^{Co/Co}$;TgMMTV-Cre;$p53^{+/-}$, $Brca1^{+/-}$; $p53^{+/-}$; IR, TgC3(1)-Tag, TgWAP-Tag). We applied a previously published p53 gene signature, which was predictive of p53 mutation status and outcome in human breast tumors, to query the mouse models in our data set. The majority of the tumors with perturbed p53 showed high expression of the average of this "p53 up" signature and low expression of the "p53 down" signature. This further validates this signature and it also validates these mouse models as mimicking p53 loss seen in human breast tumors.

As we continue adding new murine models to this data set we learn more about their biology and how they relate to one another and to human breast cancer. Our comparison of mouse models has begun to hint at the cell types of origin of these tumors, some of which seem to have potential human correlates, while others do not. Studying the cellular origins of cancers, whether the result of transformation of a stem cell, progenitor cell, or de-

differentiation of a mature cell, is a challenging task in human model systems and therefore is a major strength to be taken advantage of in mouse models.

MATERIALS AND METHODS

Mouse tumor models

Frozen tumors or total RNA was provided from the mouse models new to this study from the following investigators: MMTV-Fgf3 **(**Katrina Podsypanina and Harold Varmus - MSKCC)[20], MMTV-DN89β-catenin (Pam Cowin – NYU)[22], WAP-Cre; p53flox/flox (Eva Lee – UC Irvine)[3], $p53^{-/-}$ spontaneous (Daniel Medina – BCM)[23], $p18^{-/-}$ and $p18^{-/-}$; $Brca1^{+/-}$ (Yue Xiong – UNC), MMTV-AIB1 (Myles Brown – DFCI)[2], WAP-Cre; Tel-NTRK3 (Zhi Li and Stuart Orkin – Children's Hospital Boston), $Brg1^{+/-}$ (Scott Bultman and Terry Magnussen – UNC)[6], MMTV-HRAS (Yi Li – BCM and Katrina Podsypanina and Harold Varmus – MSKCC)[24].

RNA preparation

Total RNA was collected from mouse mammary tumors or wild type mammary glands and purified using the Qiagen RNeasy Mini Kit according to the manufacturer's protocol using 20-30 mg tissue. RNA integrity was assessed using the RNA 6000 Nano LabChip kit followed by analysis using a Bioanalyzer (Agilent).

Microarray Experiments.

2.5 micrograms of total RNA was reverse transcribed, amplified and labeled with Cy5 using a Low RNA Input Amplification kit (Agilent). The reference RNA was reverse transcribed,

amplified, and labeled with Cy3. The amplified experimental sample and reference were combined and hybridized overnight to Agilent Mouse Oligo Microarrays (G4121A or B). They were then washed and scanned on an Axon GenePix 4000B scanner, analyzed using GenePix 4.1 software and uploaded into our database where a Lowess normalization is performed.

Microarray Data Analysis

This study included the 122 microarrays previously published and available from the University of North Carolina Microarray Database (UMD and at the Gene Expression Omnibus under the series GSE3165. We also added an additional 113 microarrays. The genes for all analyses were filtered by requiring the Lowess normalized intensity values in both channels to be > 30. The $\log_2$ ratio of Cy5/Cy3 was then reported for each gene. In the final dataset, only genes that reported values in 70% or more of the samples were included. The genes were median centered and then hierarchical clustering was performed using Cluster v2.12. JavaTreeview v1.0.8 was used for cluster viewing and display. We derived a new intrinsic gene list as we had done before. Intrinsic "groups" were chosen based upon experimental samples having a Pearson correlation value of 0.65 or greater from the unsupervised clustering analysis of the expanded set of 235 murine samples. The analysis was performed using the Intrinsic Gene Identifier v1.0 developed by Max Diehn/Stanford University. Technical replicates were removed from the file and the members of every highly correlated node were given identical class numbers. Every sample that fell outside the 0.65 correlation cut off was given a class of their own. A list of 1802 "intrinsic" genes was selected using the criteria of one standard deviation below the mean intrinsic gene value.

Pathway Analysis

Two gene signatures were obtained, an ER regulated gene signature[29] and a 52 gene p53 inactivation signature [5]. Murine orthlogs were found for these genes and the data extracted. Separate lists were used for the up and down genes and the genes were standardized, median centered, and the profile was averaged for each sample.

Gene Set Enrichment Analysis.

The murine samples were classified based upon their clustering pattern in Figure 3.1 that used the new mouse intrinsic gene list, and were assigned to Groups I-XII. Two-class unpaired Significance Analysis of Microarrays analysis was performed for each murine class separately versus all other classes using an FDR of 1%. Group V did not return a gene list at FDR of 1% and therefore, we were left 11 class-specific gene lists. Using only the set of highly expressed genes that were associated with each analysis (and ignoring the genes whose low expression correlated with a given class), GSEA was performed using GSEA in R. The 11 murine gene sets were then compared to each human subtype-ranked gene sets and significant enrichments reported. For statistical strength of these enrichments, GSEA uses Family Wise Error Rate (FWER) to correct for multiple testing and False Discovery Rate to reduce false positive reporting.

REFERENCES

1. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S *et al*: **Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors**. *Genome Biol* 2007, **8**(5):R76.

2. Torres-Arzayus MI, De Mora JF, Yuan J, Vazquez F, Bronson R, Rue M, Sellers WR, Brown M: **High tumor incidence and activation of the PI3K/AKT pathway in transgenic mice define AIB1 as an oncogene**. *Cancer Cell* 2004, **6**(3):263-274.

3. Lin SC, Lee KF, Nikitin AY, Hilsenbeck SG, Cardiff RD, Li A, Kang KW, Frank SA, Lee WH, Lee EY: **Somatic mutation of p53 leads to estrogen receptor alpha-positive and -negative mouse mammary tumors with high frequency of metastasis**. *Cancer Res* 2004, **64**(10):3525-3532.

4. Zhang X, Podsypanina K, Huang S, Mohsin SK, Chamness GC, Hatsell S, Cowin P, Schiff R, Li Y: **Estrogen receptor positivity in mammary tumors of Wnt-1 transgenic mice is influenced by collaborating oncogenic mutations**. *Oncogene* 2005, **24**(26):4220-4231.

5. Troester MA, Herschkowitz JI, Oh DS, He X, Hoadley KA, Barbier CS, Perou CM: **Gene expression patterns associated with p53 status in breast cancer**. *BMC Cancer* 2006, **6**:276.

6. Bultman SJ, Herschkowitz JI, Godfrey V, Gebuhr TC, Yaniv M, Perou CM, Magnuson T: **Characterization of mammary tumors from Brg1 heterozygous mice**. *Oncogene* 2007.

7. Sandgren EP, Schroeder JA, Qui TH, Palmiter RD, Brinster RL, Lee DC: **Inhibition of mammary gland involution is associated with transforming growth factor alpha but not c-myc-induced tumorigenesis in transgenic mice**. *Cancer Res* 1995, **55**(17):3915-3927.
8. Gallahan D, Jhappan C, Robinson G, Hennighausen L, Sharp R, Kordon E, Callahan R, Merlino G, Smith GH: **Expression of a truncated Int3 gene in developing secretory mammary epithelium specifically retards lobular differentiation resulting in tumorigenesis**. *Cancer Res* 1996, **56**(8):1775-1785.

9. Simin K, Wu H, Lu L, Pinkel D, Albertson D, Cardiff RD, Van Dyke T: **pRb inactivation in mammary cells reveals common mechanisms for tumor initiation and progression in divergent epithelia**. *PLoS Biol* 2004, **2**(2):E22.

10. Husler MR, Kotopoulis KA, Sundberg JP, Tennent BJ, Kunig SV, Knowles BB: **Lactation-induced WAP-SV40 Tag transgene expression in C57BL/6J mice leads to mammary carcinoma**. *Transgenic Res* 1998, **7**(4):253-263.

11. Maroulakou IG, Anver M, Garrett L, Green JE: **Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene**. *Proc Natl Acad Sci U S A* 1994, **91**(23):11236-11240.

12. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ: **Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease**. *Proc Natl Acad Sci U S A* 1992, **89**(22):10578-10582.

13. Tsukamoto AS, Grosschedl R, Guzman RC, Parslow T, Varmus HE: **Expression of the int-1 gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice**. *Cell* 1988, **55**(4):619-625.

14. Guy CT, Cardiff RD, Muller WJ: **Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease**. *Mol Cell Biol* 1992, **12**(3):954-961.

15. Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, Hennighausen L, Wynshaw-Boris A, Deng CX: **Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation**. *Nat Genet* 1999, **22**(1):37-43.

16. Jerry DJ, Kittrell FS, Kuperwasser C, Laucirica R, Dickinson ES, Bonilla PJ, Butel JS, Medina D: **A mammary-specific model demonstrates the role of the p53 tumor suppressor gene in tumor development**. *Oncogene* 2000, **19**(8):1052-1058.

17. Yin Y, Bai R, Russell RG, Beildeck ME, Xie Z, Kopelovich L, Glazer RI: **Characterization of medroxyprogesterone and DMBA-induced multilineage mammary tumors by gene expression profiling**. *Mol Carcinog* 2005.

18. Backlund MG, Trasti SL, Backlund DC, Cressman VL, Godfrey V, Koller BH: **Impact of ionizing radiation and genetic background on mammary tumorigenesis in p53-deficient mice**. *Cancer Res* 2001, **61**(17):6577-6582.

19. Cressman VL, Backlund DC, Hicks EM, Gowen LC, Godfrey V, Koller BH: **Mammary tumor formation in p53- and BRCA1-deficient mice**. *Cell Growth Differ* 1999, **10**(1):1-10.

20. Kwan H, Pecenka V, Tsukamoto A, Parslow TG, Guzman R, Lin TP, Muller WJ, Lee FS, Leder P, Varmus HE: **Transgenes expressing the Wnt-1 and int-2 proto-oncogenes cooperate during mammary carcinogenesis in doubly transgenic mice**. *Mol Cell Biol* 1992, **12**(1):147-154.

21. Muller WJ, Lee FS, Dickson C, Peters G, Pattengale P, Leder P: **The int-2 gene product acts as an epithelial growth factor in transgenic mice**. *Embo J* 1990, **9**(3):907-913.

22. Imbert A, Eelkema R, Jordan S, Feiner H, Cowin P: **Delta N89 beta-catenin induces precocious development, differentiation, and neoplasia in mammary gland**. *J Cell Biol* 2001, **153**(3):555-568.

23. Kuperwasser C, Hurlbut GD, Kittrell FS, Dickinson ES, Laucirica R, Medina D, Naber SP, Jerry DJ: **Development of spontaneous mammary tumors in BALB/c p53 heterozygous mice. A model for Li-Fraumeni syndrome**. *Am J Pathol* 2000, **157**(6):2151-2159.

24. Sinn E, Muller W, Pattengale P, Tepler I, Wallace R, Leder P: **Coexpression of MMTV/v-Ha-ras and MMTV/c-myc genes in transgenic mice: synergistic action of oncogenes in vivo**. *Cell* 1987, **49**(4):465-475.

25. Li Y, Welm B, Podsypanina K, Huang S, Chamorro M, Zhang X, Rowlands T, Egeblad M, Cowin P, Werb Z *et al*: **Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells**. *Proc Natl Acad Sci U S A* 2003, **100**(26):15853-15858.

26. Rosner A, Miyoshi K, Landesman-Bollag E, Xu X, Seldin DC, Moser AR, MacLeod CL, Shyamala G, Gillgrass AE, Cardiff RD: **Pathway pathology: histological differences between ErbB/Ras and Wnt pathway transgenic mammary tumors**. *Am J Pathol* 2002, **161**(3):1087-1097.

27. Shackleford GM, MacArthur CA, Kwan HC, Varmus HE: **Mouse mammary tumor virus infection accelerates mammary carcinogenesis in Wnt-1 transgenic mice by insertional activation of int-2/Fgf-3 and hst/Fgf-4**. *Proc Natl Acad Sci U S A* 1993, **90**(2):740-744.

28. Desai KV, Xiao N, Wang W, Gangi L, Greene J, Powell JI, Dickson R, Furth P, Hunter K, Kucherlapati R *et al*: **Initiating oncogenic event determines gene-expression patterns of human breast cancer models**. *Proc Natl Acad Sci U S A* 2002, **99**(10):6967-6972.

29. Oh DS, Troester MA, Usary J, Hu Z, He X, Fan C, Wu J, Carey LA, Perou CM: **Estrogen-Regulated Genes Predict Survival in Hormone Receptor-Positive Breast Cancers**. *J Clin Oncol* 2006.

30. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc Natl Acad Sci U S A* 2001, **98**(19):10869-10874.

CHAPTER IV

PHENOTYPIC EVALUATION OF A NEW MOLECULAR SUBTYPE OF HUMAN INVASIVE BREAST CARCINOMA

ABSTRACT

Global gene expression profiling of human breast cancers has identified five distinct and prevalent subtypes of tumors (luminal A, luminal B, normal breast-like, HER2+/ER-, and basal-like). As a result of continuing these microarray studies we have identified a new and relatively infrequent (3-5%) human subtype that was not apparent in smaller datasets. This subtype, referred to as "Claudin-low", is defined by the low expression of genes involved in tight junctions and cell-cell adhesion including Claudins 3 and 12, Occludin, and E-cadherin. These tumors also show a marked increased expression of lymphocyte markers, a low expression of luminal genes, inconsistent basal gene expression, and high expression of endothelial cell markers. The aim of this study was to characterize the histologic and immunophenotypic properties of "Claudin-low" tumors that were first positively identified using DNA microarray analysis. Detailed histologic review and immunohistochemistry for Claudin 3 and E-cadherin was performed on tumors with known microarray profiles including 19 "Claudin-low" tumors. Most of these tumors were described as high grade ductal carcinomas with marked stromal desmoplasia. Several "Claudin-low" tumors showed

increased tumor infiltrating lymphocytes and most were ER- and HER2-negative. Consistent with their transcriptional profiles, immunohistochemical analysis showed that most claudin-low tumors expressed Claudin 3 and E-cadherin at low levels.

INTRODUCTION

Five distinct and prevalent subtypes of human breast tumors (luminal A, luminal B, normal breast-like, HER2+/ER-, and basal-like) have been identified using expression profiling[1-3]. As a result of continuing these microarray studies with larger patient sets, it is anticipated that there will be further stratification of these subtypes in addition to uncovering new molecular subtypes that were not appreciated before because of their infrequent occurrence. We believe we have identified a new and relatively rare (3-5%) human subtype that was not apparent in smaller datasets[4]. This subtype, referred to as "Claudin-low", is defined by a gene expression pattern that includes low levels of genes involved in tight junctions and cell-cell adhesion including Claudins 3 and 12, Occludin, and E-cadherin. These tumors show a marked increased expression of B cell and T cell markers. These tumors also showed low expression of luminal epithelial genes. The aim of this study was to characterize the histologic and immunophenotypic properties of "Claudin-low" tumors that were first positively identified using DNA microarray analysis. Detailed histologic review was performed on tumors with known microarray profiles including 19 "Claudin-low" tumors and immunohistochemistry for Claudin 3 and E-cadherin was performed.

RESULTS

A new intrinsic subtype shows low expression of tight junction and cell-cell adhesion associated genes

We clustered a breast cancer dataset generated at UNC, containing 482 microarrays consisting of 349 primary breast tumor samples, 3 breast tumor xenografts, and 18 normal breast samples assayed on Agilent microarrays using the previously published intrinsic gene set[1] (Figure 4.1). As before, this clustering separated the samples into distinct subtypes (luminal A+B combined, normal breast-like, HER2+/ER-, basal-like, and basal/HER2). This clustering also revealed an additional subtype which shows low expression of a cluster of genes including the tight junction and cell-cell adhesion associated genes Claudin 3, Occludin, and E-cadherin. We therefore referred to this new subtype as Claudin-low. Claudin-low tumors were also low for the expression of luminal tumor defining genes and inconsistent in expression for the basal-like gene cluster.

Figure 4.1

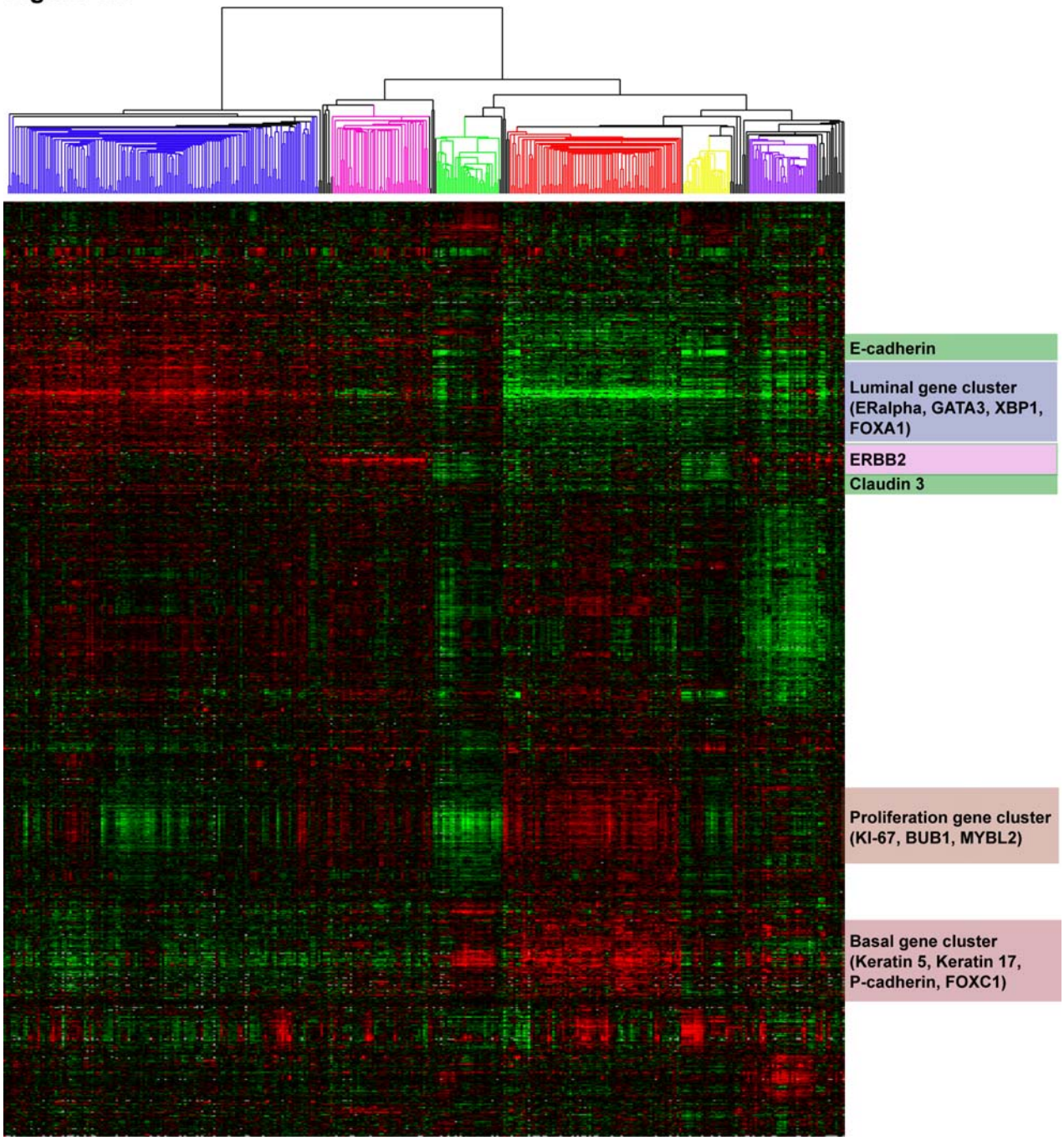

E-cadherin
Luminal gene cluster (ERalpha, GATA3, XBP1, FOXA1)
ERBB2
Claudin 3
Proliferation gene cluster (KI-67, BUB1, MYBL2)
Basal gene cluster (Keratin 5, Keratin 17, P-cadherin, FOXC1)

Figure 4.1. Hierchical clustering of breast tumors using intrinsic gene list. The dendrogram is colored to represent the intrinsic subtypes; luminal=blue, HER2/ER-=pink, normal-like=green, basal-like=red, basal/HER2=purple and the claudin-low subtype=yellow.

Morphological and clinical features of Claudin-low tumors

We performed a detailed histological review of the Claudin-low tumors (Figure 4.2 and Appendix IIC). These tumors predominantly included high grade ductal carcinomas that were associated with marked stromal desmoplasia, which is the growth of dense fibrous or connective tissue around the tumor (Figure 4.2a and b). Several of these tumors show increased tumor infiltrating lymphocytes (Figure 4.2b and c). These tumors were not reported to be inflammatory breast cancer (IBC). IBC is caused by blockage of the dermal lymphatics by infiltration of tumor cells, rather than an infiltration of inflammatory cells. In addition, unlike Claudin-low tumors, IBC tumors have not been shown to produce high levels of inflammatory cytokines and inflammatory cells are only rarely detected around tumor stroma. These Claudin-low tumors are not invasive lobular carcinomas (ILC), which are also breast tumors that show low levels of E-cadherin. In ILC, E-cadherin is typically low due to the result of an E-cadherin mutation. ILC also have a distinct morphology and are usually ER-positive, whereas Claudin-low tumors are predominantly clinically ER and HER2 negative (Table 4.1) and therefore would be defined as triple negative.

Figure 4.2

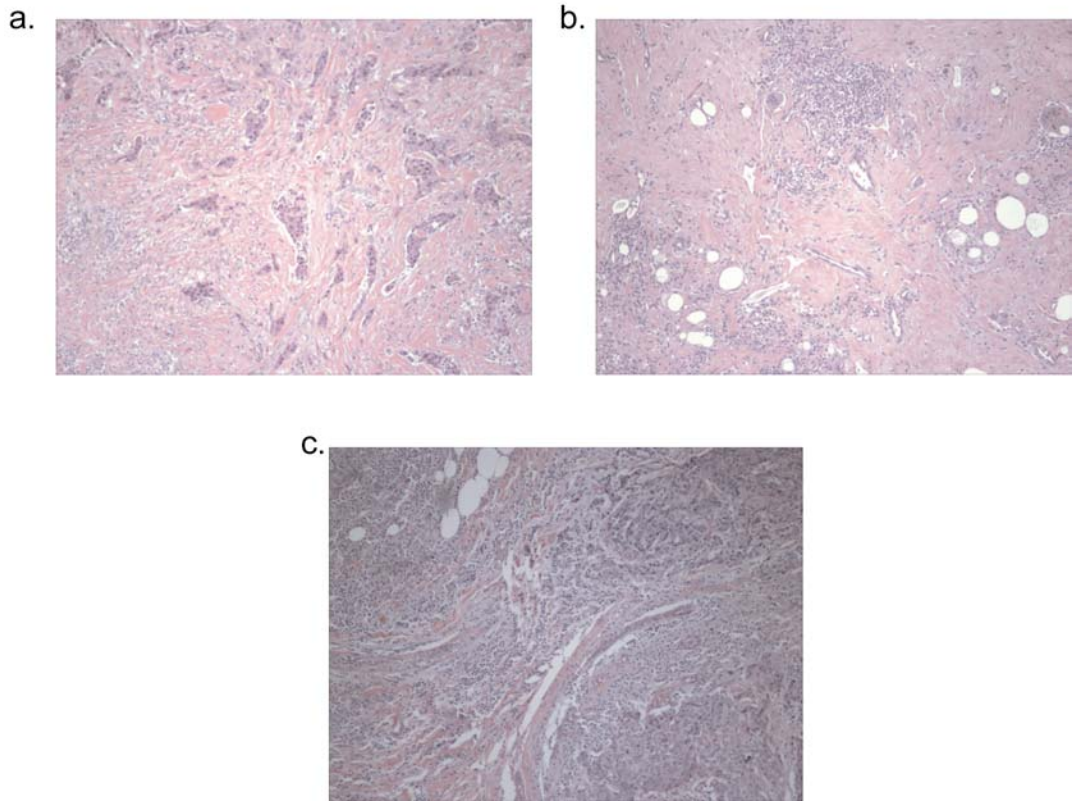


Figure 4.2. Morpological features of Claudin-low tumors. Claudin-low tumors are mostly high grade ductal carcinomas associated with marked stromal desmoplasia (a,b). Several of these tumors show increased tumor infiltrating lymphocytes (b,c).

Table 4.1. Claudin-low tumors are predominantly clinically ER and HER2 negative.

	ER	HER2
1	-	N/A
2	-	N/A
3	-	-
4	-	-
5	-	-
6	-	+
7	-	-
8	-	N/A
9	-	N/A
10	-	+
11	N/A	N/A
12	N/A	N/A
13	-	-
14	-	-
15	-	-
16	+	N/A
17	-	-
18	+	-

Gene expression signature associated with Claudin-low breast carcinomas

To determine the gene expression signature associated with the Claudin-low subtype, a 2-class Significance Analysis of Microarrays (SAM) was performed[5]. In total there were 1704 genes identified with a false discovery rate of 1%. 1107 genes were significantly highly expressed regulated in Claudin-low tumors. An analysis of Gene Ontology showed that categories of immune and inflammatory response were the top biological processes with Bonferroni-corrected significant enrichment scores in this gene list (Table 4.2). These tumors show a marked increased expression of B cell markers (CD19) and T cell markers (CD3z and CD3g). One of the top genes in the SAM list was the general lymphocyte marker, CD45, which again suggests the presence of immune cell infiltrates in Claudin-low tumors.

Table 4.2a. Gene ontology categories positively enriched in Claudin-low tumors (top25).

System	Gene Category	List Hits	List Total	Population Hits	Population Total	EASE score	Bonferroni
GO Biological Process	RESPONSE TO BIOTIC STIMULUS	170	797	873	10997	3.81E-35	9.31E-32
GO Biological Process	IMMUNE RESPONSE	154	797	741	10997	5.02E-35	1.23E-31
GO Biological Process	DEFENSE RESPONSE	164	797	834	10997	2.40E-34	5.86E-31
GO Biological Process	RESPONSE TO PEST, PATHOGEN OR PARASITE	99	797	513	10997	1.05E-19	2.57E-16
GO Biological Process	RESPONSE TO STIMULUS	231	797	1830	10997	1.84E-19	4.50E-16
GO Biological Process	RESPONSE TO OTHER ORGANISM	100	797	539	10997	1.21E-18	2.95E-15
GO Biological Process	ORGANISMAL PHYSIOLOGICAL PROCESS	210	797	1702	10997	2.52E-16	6.16E-13
GO Molecular Function	SIGNAL TRANSDUCER ACTIVITY	240	830	2232	11942	1.16E-13	2.83E-10
GO Biological Process	RESPONSE TO EXTERNAL STIMULUS	80	797	474	10997	1.13E-12	2.76E-09
GO Biological Process	HUMORAL IMMUNE RESPONSE	40	797	154	10997	2.34E-12	5.70E-09
GO Biological Process	RESPONSE TO STRESS	129	797	970	10997	4.68E-12	1.14E-08
GO Biological Process	INFLAMMATORY RESPONSE	45	797	206	10997	4.59E-11	1.12E-07
GO Biological Process	SIGNAL TRANSDUCTION	263	797	2549	10997	4.80E-11	1.17E-07
GO Cellular Component	PLASMA MEMBRANE	180	761	1591	10532	1.23E-10	3.00E-07
GO Biological Process	RESPONSE TO WOUNDING	62	797	363	10997	3.52E-10	8.60E-07
GO Biological Process	HUMORAL DEFENSE MECHANISM (SENSU VERTEBRATA)	29	797	111	10997	3.31E-09	8.10E-06
GO Cellular Component	IMMUNOLOGICAL SYNAPSE	15	761	30	10532	4.52E-09	1.10E-05
GO Biological Process	CELL COMMUNICATION	273	797	2799	10997	7.69E-09	1.88E-05
GO Biological Process	IMMUNE CELL ACTIVATION	26	797	103	10997	5.27E-08	1.29E-04
GO Cellular Component	LYSOSOME	28	761	118	10532	5.47E-08	1.34E-04
GO Cellular Component	LYTIC VACUOLE	28	761	118	10532	5.47E-08	1.34E-04
GO Biological Process	CELL ACTIVATION	26	797	104	10997	6.48E-08	1.58E-04
GO Biological Process	LYMPHOCYTE ACTIVATION	23	797	85	10997	9.75E-08	2.38E-04
GO Cellular Component	VACUOLE	29	761	133	10532	2.04E-07	4.98E-04
GO Molecular Function	RECEPTOR ACTIVITY	143	830	1362	11942	2.37E-07	5.80E-04

Table 4.2b. Gene ontology categories negatively enriched in Claudin-low tumors.

System	Gene Category	List Hits	List Total	Population Hits	Population Total	EASE score	Bonferroni
GO Biological Process	CHROMATIN ASSEMBLY	17	370	99	10997	1.66E-07	2.44E-04
GO Biological Process	NUCLEOSOME ASSEMBLY	16	370	90	10997	2.68E-07	3.94E-04
GO Cellular Component	NUCLEOSOME	15	359	80	10532	4.20E-07	6.16E-04
GO Biological Process	CHROMATIN ASSEMBLY OR DISASSEMBLY	17	370	134	10997	1.03E-05	1.51E-02
GO Biological Process	PROTEIN COMPLEX ASSEMBLY	25	370	274	10997	1.67E-05	2.45E-02
GO Cellular Component	CHROMATIN	17	359	143	10532	2.72E-05	4.00E-02

Claudin-low tumors also express Snail, Slug, and Twist 2; transcription factors (or family members thereof in the case of Twist 2) that have been shown to repress the transcription of E-cadherin and that are involved in the process of epithelial-to-mesenchymal transition (EMT)[6-8]. Reduced E-cadherin expression can also be associated with abnormal expression of P-cadherin, N-cadherin, or Cadherin-11. This has been coined the "cadherin switch". Neither N-cadherin nor P-cadherin, which is high in basal-like tumors, were highly expressed in the Claudin-low subtype. Interestingly, Cadherin-11 was expressed at higher levels in Claudin-low tumors.

597 genes were significantly negatively regulated in Claudin-low tumors including tight junction associated genes (Claudin 3, E-cadherin, Occludin, and tight junction protein 2). Significant ontology categories enriched in this list included genes involved in chromatin and nucleosome assembly (Table 4.2). As we noticed from the intrinsic cluster, Claudin-low tumors also showed low expression of luminal epithelial signature genes (Keratin 8, ER, XBP1, FOXA1, and GATA3).

Claudin-low tumors express low levels of tight junction associated proteins by IHC

In order to validate these gene expression results, we performed IHC for Claudin 3 and E-cadherin on 15 basal-like, 32 luminal A, 14 luminal B, 13 HER2+/ER- and 19 claudin-low tumors. Over half of the Claudin-low tumors either showed no staining (3/19) or weak positive (1+) staining for Claudin 3 (8/19) (Figure 4.3, Table 4.3a). No tumors from any of the other subtypes showed a complete lack of Claudin 3 staining. The majority of tumors from each non-Claudin-low subtype, Basal-like (13/15), Luminal A (21/32), Luminal B

(13/14), Normal-like (7/8), and HER2+, ER- (8/13) scored either moderate or strong positive (2 or 3+) staining results. The results were similar for E-cadherin staining (Figure 4.3, Table 4.3b) where more than half of the Claudin-low tumors (10/19) were scored 0 or 1+. By comparison, Basal-like (13/15), Luminal A (24/32), Luminal B (12/14), Normal-like (8/8), and HER2+, ER- (12/13) were scored either moderate or strong positive (2 or 3+). These results were consistent, however, but not as dramatic as the gene expression data. A possible explanation is that tumors show heterogeneity for loss of expression of these proteins and detection depends on the section that is taken for staining. It also should be noted that while the low expression of Claudin 3 and E-cadherin are distinct, intrinsic, features of the Claudin-low subtype, these characteristics are not unique to Claudin-low tumors and appear to be characteristic of subsets of tumors in other subtypes as well (Figure 4.1).

Figure 4.3

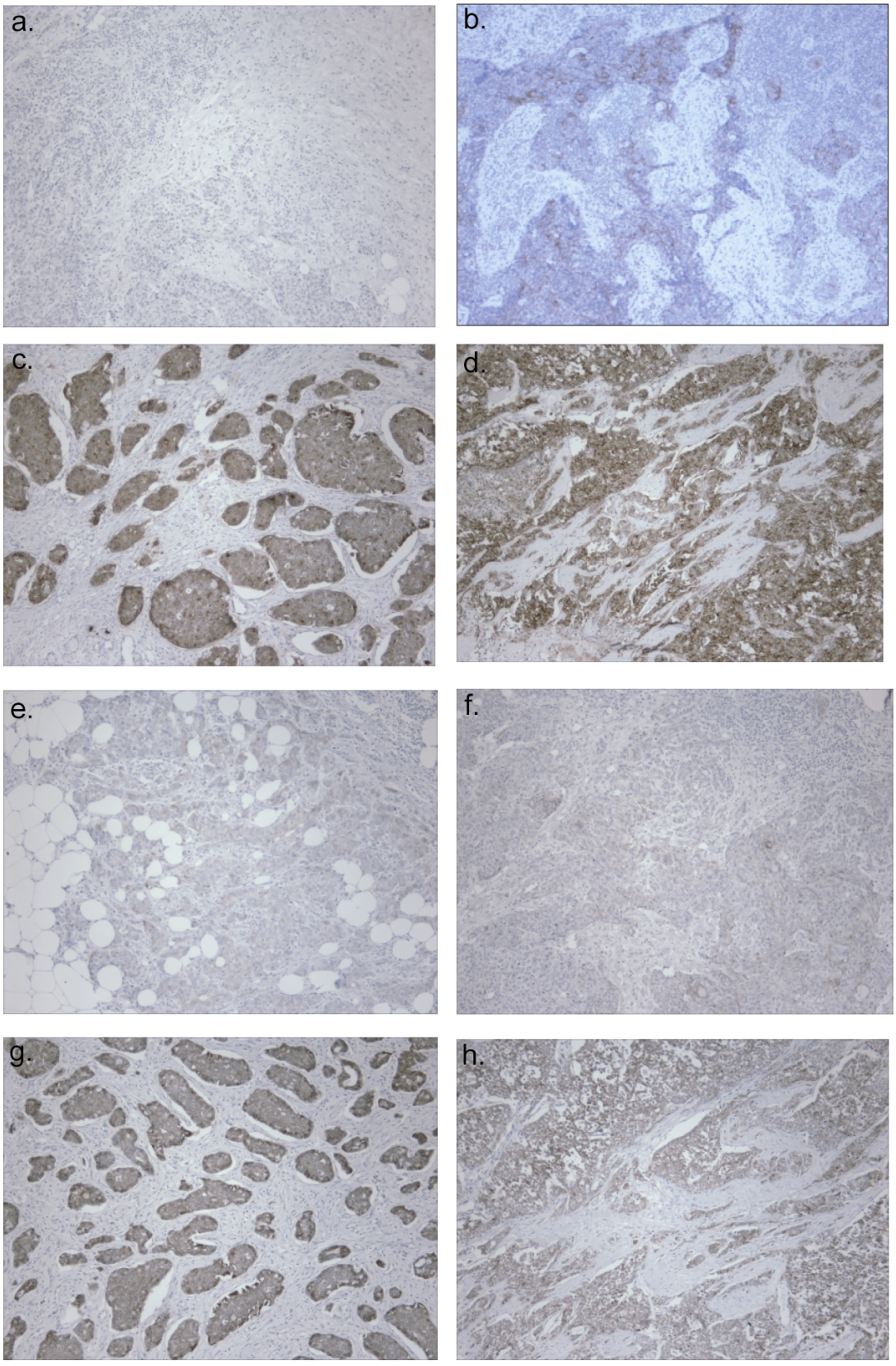


Figure 4.3. Immunophenotype of Claudin-low tumors. Claudin-low tumors (a,b,e,f) often show lower expression of Claudin 3 and E-cadherin than basal-like (d,h) and luminal tumors(c,g).

Table 4.3a. Summary of immunohistochemistry results for Claudin 3.

Subtype	0	1	2	3
Basal-like	0/15	2/15	6/15	7/15
Luminal A	0/32	11/32	13/32	8/32
Luminal B	0/14	1/14	6/14	7/14
Normal-like	0/8	1/8	5/8	2/8
HER2+, ER-	0/13	5/13	3/13	5/13
Claudin-low	3/19	8/19	7/19	1/19

Table 4.3b. Summary of immunohistochemistry results for E-cadherin.

Subtype	0	1	2	3
Basal-like	0/15	2/15	6/15	7/15
Luminal A	1/32	7/32	5/32	19/32
Luminal B	1/14	1/14	3/14	9/14
Normal-like	0/8	0/8	5/8	3/8
HER2+, ER-	0/13	1/13	5/13	7/13
Claudin-low	1/19	9/19	7/19	2/19

(0=negative, 1=weak positive, 2=moderate positive, 3=strong positive)

DISCUSSION

In this report, we present the identification and characterization of a new molecular subtype of breast cancer identified by intrinsic clustering. This subtype shows a low expression of tight junction associated genes. These tumors show high expression of immune genes and lymphocyte markers, which are features that were confirmed by the histology that shows the presence of many lymphocytic infiltrates. Loss of E-cadherin, usually by mutation and accompanying loss of heterozygosity, is a very common feature of invasive lobular carcinomas (ILC)[9]. Histology and estrogen receptor status rule out the Claudin-low tumors in this study as being ILC. Complete loss of E-cadherin has been

reported in IDC[9]. Mutations are not seen in IDC, however, E-cadherin LOH is a relatively common event.

E-cadherin expression can be regulated at the transcriptional level by repressors including Snail, Slug (Snail2), and Twist[6-8]. Snail, Slug, and Twist 2 are all highly expressed in Claudin-low tumors. Snail expression has been shown to correlate with histological grade and lymph node status in breast carcinomas[6]. It was also noted in this study that some dedifferentiated tumors also expressed Snail and were node negative. Snail has also been shown to promote tumor recurrence in a mouse mammary tumor model[10]. Concurrent with losing E-cadherin, Claudin-low tumors appear to gain expression of Cadherin-11. Normally, Cadherin-11 is found preferentially in mesenchymal tissues, but it has been detected in invasive breast carcinoma cell lines[11]. Cadherin-11 has also been shown to promote tumor cell motility and invasiveness and there has been a reported correlation between expression of Cadherin-11 and enhanced metastastic potential[12]. The presence of a Cadherin-11 splice variant has been suggested to promote invasion[13]. The expression of this gene thus needs to be validated by IHC to determine if in fact high protein expression is a feature of the tumor cells or limited to the tumor associated stroma.

The Claudin-low subtype makes up only about 3-5% of breast cancers, however, if almost 200,000 new cases arise each year in the United States, then that equals 6-12,000, which is a significant number of patients. In this study we determined and validated using IHC some of the genes/proteins that showed low expression in these tumors. Due to the high expression of genes presumably coming from infiltrative lymphocytes in these tumors, it is a

challenge to find positive IHC markers to identify this tumor class. These finding when and if they occur, should assist in the prospective and retrospective identification of Claudin-low breast tumor clinical specimens, which will facilitate further molecular, epidemiologic, and treatment studies of this rare and unique tumor subtype.

MATERIALS AND METHODS

Significance analysis of microarrays (SAM) and gene ontology analysis

SAM analysis was performed to identify genes that were significantly different between Claudin-low tumors compared with the remaining samples[5]. Expression analysis systemic explorer (EASE) was used to identify gene ontology categories overrepresented in the Claudin-low gene list compared to the genes present on the array.

Immunohistochemistry

Formalin-fixed, paraffin-embedded tissue sections (~5um) were processed using standard immunostaining methods. Following deparaffinization in xylenes, slides were rehydrated through a graded series of alcohol and rinsed in phosphate buffered saline. Endogenous peroxidase activity was blocked with 3% hydrogen peroxidase. Samples were steamed for antigen retrieval with 10 mM citrate buffer (pH 6.0) for 30 min. Slides were then incubated for 20 minutes with diluted normal blocking serum. The sections were incubated for 60 minutes at room temperature with primary antibody to claudin 3 (Zymed 18-7340 1:100) or e-cadherin (Cell Marque ECH-6 pre-diluted) . The slides were incubated for 45 minutes with diluted biotinylated secondary antibody (1:250 dilution) and 30 minutes with Vectastain Elite

ABC reagent (Vector Laboratories). Sections were incubated in peroxidase substrate solution for visualization.. Slides were counterstained with hematoxylin and examined by light microscopy. Tumor immunoreactivity was scored 0=negative, 1=weak positive, 2=moderate positive, and 3=strong positive.

REFERENCES

1. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L *et al*: **The molecular portraits of breast tumors are conserved across microarray platforms**. *BMC Genomics* 2006, **7**(1):96.

2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc Natl Acad Sci U S A* 2001, **98**(19):10869-10874.

3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: **Repeated observation of breast tumor subtypes in independent gene expression data sets**. *Proc Natl Acad Sci U S A* 2003, **100**(14):8418-8423.

4. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S *et al*: **Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors**. *Genome Biol* 2007, **8**(5):R76.

5. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response**. *Proc Natl Acad Sci U S A* 2001, **98**(9):5116-5121.

6. Blanco MJ, Moreno-Bueno G, Sarrio D, Locascio A, Cano A, Palacios J, Nieto MA: **Correlation of Snail expression with histological grade and lymph node status in breast carcinomas**. *Oncogene* 2002, **21**(20):3241-3246.

7. Bolos V, Peinado H, Perez-Moreno MA, Fraga MF, Esteller M, Cano A: **The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors**. *J Cell Sci* 2003, **116**(Pt 3):499-511.

8. Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C, Savagner P, Gitelman I, Richardson A, Weinberg RA: **Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis**. *Cell* 2004, **117**(7):927-939.

9. Asgeirsson KS, Jonasson JG, Tryggvadottir L, Olafsdottir K, Sigurgeirsdottir JR, Ingvarsson S, Ogmundsdottir HM: **Altered expression of E-cadherin in breast cancer. patterns, mechanisms and clinical significance**. *Eur J Cancer* 2000, **36**(9):1098-1106.

10. Moody SE, Perez D, Pan TC, Sarkisian CJ, Portocarrero CP, Sterner CJ, Notorfrancesco KL, Cardiff RD, Chodosh LA: **The transcriptional repressor Snail promotes mammary tumor recurrence**. *Cancer Cell* 2005, **8**(3):197-209.

11. Pishvaian MJ, Feltes CM, Thompson P, Bussemakers MJ, Schalken JA, Byers SW: **Cadherin-11 is expressed in invasive breast cancer cell lines**. *Cancer Res* 1999, **59**(4):947-952.

12. Nieman MT, Prudoff RS, Johnson KR, Wheelock MJ: **N-cadherin promotes motility in human breast cancer cells regardless of their E-cadherin expression**. *J Cell Biol* 1999, **147**(3):631-644.

13. Feltes CM, Kudo A, Blaschuk O, Byers SW: **An alternatively spliced cadherin-11 enhances human breast cancer cell invasion**. *Cancer Res* 2002, **62**(22):6688-6697.

CHAPTER V

LOSS OF THE RETINOBLASTOMA TUMOR SUPPRESSOR IS A COMMON EVENT IN BASAL-LIKE AND LUMINAL B BREAST CARCINOMAS

PREFACE

This chapter represents a manuscript that is being prepared for submission. I performed data analysis, scoring of IHC staining, development of the figures, and the writing of the manuscript. Xiaping He prepared DNA from tumors and normal patient lymphocytes, and IHC staining and scoring. Cheng Fan assisted with the data analysis. Charles Perou was the Principal Investigator, conceived and designed the study, and helped draft the paper.

Jason I Herschkowitz, Xiaping He, Cheng Fan, and Charles M. Perou. (2007) Loss of the retinoblastoma tumor suppressor is a common event in Basal-like and Luminal B breast carcinomas. [in preparation]

SUMMARY

Breast cancers can be classified using whole genome expression into distinct subtypes that show differences in patient prognosis. One of these groups, the basal-like carcinomas, are poorly differentiated, highly metastatic, and genomically unstable. These tumors also contain specific genetic alterations with one example being frequent p53 mutations. The loss of the tumor suppressor gene encoded by the retinoblastoma (RB1) locus is a well-characterized occurrence in many tumor types. However, its role in breast cancer is less clear with many reports demonstrating a Loss of Heterozygosity (LOH), but which does not correlate with loss of RB1 protein expression. Here we report that LOH of the RB1 locus was observed at a high frequency in basal-like and luminal B tumors. These tumors also concurrently have low expression of RB1 mRNA as assessed by DNA microarray. As in previous reports, we did not see a significant correlation between RB1 LOH and protein expression as measured by immunohistochemistry (IHC). $p16^{INK4a}$, however, was highly expressed both by microarray and IHC, in basal-like tumors presumably due to a previously reported feedback loop caused by RB1 loss. These results suggest that the functional loss of RB1 is a common event in the progression of basal-like and luminal B breast tumors, which may play a key role in dictating therapeutic responses.

SIGNIFICANCE

Gene expression profiling has separated breast tumors into clinically prognostic subtypes, however, many of the underlying causative mutations are still not known. Here we report that LOH at the RB1 locus was observed at a high frequency in two subtypes (basal-like and luminal B tumors), both of which are highly proliferative. In addition, in basal-like

tumors, the RB1 LOH is also accompanied by the high expression of p16. The loss of RB1 is likely to play a key role in the distinct poor prognosis and response to treatment that is characteristic these subtypes.

INTRODUCTION

The retinoblastoma tumor suppressor gene (RB1) encodes a nuclear phosphoprotein that plays a central role in the regulation of the cell cycle. Inactivation of both alleles of this gene is involved in the development of retinoblastoma, which is a rare childhood malignancy. The loss of RB1 is also a well-characterized occurrence in many other human tumor types and it is likely that the p16-RB pathway is disrupted in most human tumors[1]. RB1 regulates progression through the G1 to S-phase transition of the cell cycle. In cells entering the cell cycle, extracellular signals induce the expression of D-type cyclins, which bind to and activate cyclin-dependent kinases (CDK4 and CDK6); these complexes in turn lead to the phosphorylation of RB, its dissociation from E2F family members that then transcriptionally activate many genes required for S phase. The INK4 family of cyclin-dependent kinase inhibitors ($p16^{INK4a}$, $p15^{INK4b}$, $p18^{INK4c}$, and $p19^{INK4d}$) inhibit CDK4 and CDK6 retaining RB in its hypo-phosphorylated E2F associated state, therefore preventing G1 to S-phase progression. It has recently been shown that CDK4 and CDK6 (and CDK2) are dispensable for driving the essential cell cycle, however they are required in specialized tissues and possibly to achieve higher levels of proliferation[2].

Inactivation of the RB1 gene in breast cancer was originally shown using a series of cell lines[3]. Subsequently, loss of heterozygosity (LOH) has been observed in primary tumors, but does not necessarily correlate with loss of RB1 protein expression by IHC[4, 5]. LOH has, however, been shown to correlate with RB1 mRNA low expression [4]. There are also events upstream of RB1 that may be present in breast tumors, which can negatively impact RB function by promoting its phosphorylation that include p16INK4a loss[6] and cyclin D1 amplification/overexpression[7]

Breast cancer is a heterogeneous disease, which is separable into clinically significant subtypes as defined by molecular profiling[8, 9]. In addition to reproducible gene expression differences between these subtypes, specific molecular alterations continue to be identified that correlate with each subtype. Tumors of the basal-like subtype generally have a high mitotic index, tend to be p53 mutated [10] and highly express the proliferation signature, which is a gene cluster shown to contain many E2F target genes[11, 12]. Here we report that LOH at the RB1 locus occurs at a high frequency in basal-like and Luminal B tumors, while occurring infrequently in Luminal A and HER2+/ER- tumors. Many of these tumors also concurrently have low expression of RB1 mRNA as assessed by DNA microarray. p16INK4A is also highly expressed both by microarray and by IHC in most of these RB1 LOH basal-like tumors presumably due to a feedback caused by RB1 loss. These results further illustrate the unique biology of the basal-like subtype.

RESULTS

Basal-like tumors show low expression of the RB1 transcript

Tumors of the basal-like subtype have been shown in several studies to have a high mitotic rate and to highly express a proliferation gene signature[10, 11]. For this reason, we postulated that there might be a defect in the RB pathway in these tumors. We first examined the expression levels of the core components and regulators of the RB pathway in a previously published microarray data set[8, 13] that contains 232 microarrays consisting of 184 primary breast tumor samples and 9 normal breast samples (Figure 5.1). On average, RB1 was expressed at the lowest levels in basal-like tumors and at the highest levels in luminal A tumors (Figure 5.1 and 5.2A). Tumors of the basal-like subtype also frequently showed high levels of p16 INK4A (Figure 5.1 and 5.2B) and E2F1 (Figure 5.1 and not shown). Cyclin D1 levels, on the other hand, were elevated in mainly luminal tumors including noticeably higher expression in many Luminal B tumors (Figure 5.1 and 5.2C); in addition, it was recently reported that there are often high level gains of the 11q13 locus that includes Cyclin D1 in luminal tumors (A and B)[14], thus suggesting that Cyclin D1 high expression can be considered a "luminal event".

LOH at the RB1 locus is a common event in breast cancer associated with high proliferation

Due to the low expression of RB1 message in basal-like tumors we decided to examine these breast carcinomas for LOH at the RB1 locus. We investigated 88 paired primary human breast carcinomas and normal tissue genomic DNA samples to assess the

frequency of LOH in RB1. We used two polymorphic markers located at the RB1 locus (13q14); a variable number tandem repeat (VNTR) in intron 20 and D13S153, a microsatellite marker located within intron 2. There were 67 cases that were informative for at least one of these two markers. In total, 26 tumors showed RB1 LOH for at least one marker (26/67 38.8%). This is consistent with the frequency of RB1 LOH seen in previous studies of breast cancer (26-47%)[4, 5, 15].

Figure 5.1

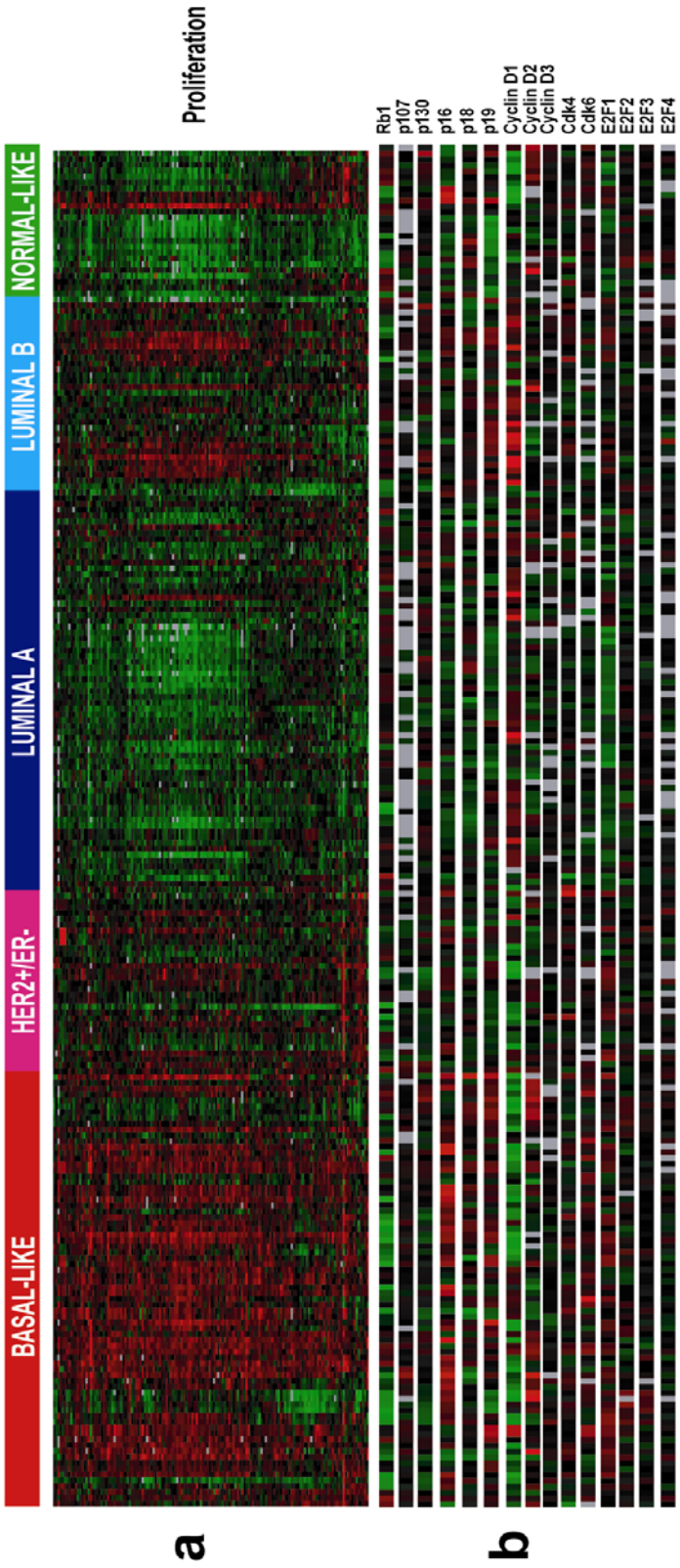

BASAL-LIKE
HER2+/ER-
LUMINAL A
LUMINAL B
NORMAL-LIKE
relative to median expression
>6
>4
>2
1:1
>2
>4
>6
Proliferation
Rb1
p107
p130
p16
p18
p19
Cyclin D1
Cyclin D2
Cyclin D3
Cdk4
Cdk6
E2F1
E2F2
E2F3
E2F4
a
b

Figure 5.1 The expression of retinoblastoma pathway members varies across breast cancer intrinsic subtypes. 232 human samples are ordered by subtype according to the 5-class single sample predictor from Hu et al. 2006[8]. Samples are colored according to their subtype: red=Basal-like, dark blue=Luminal A, light blue=Luminal B, pink= HER2+/ER-, and green=Normal breast-like. a) proliferation gene cluster b) Rb pathway genes which are present on array and pass filtering criteria.

Figure 5.2

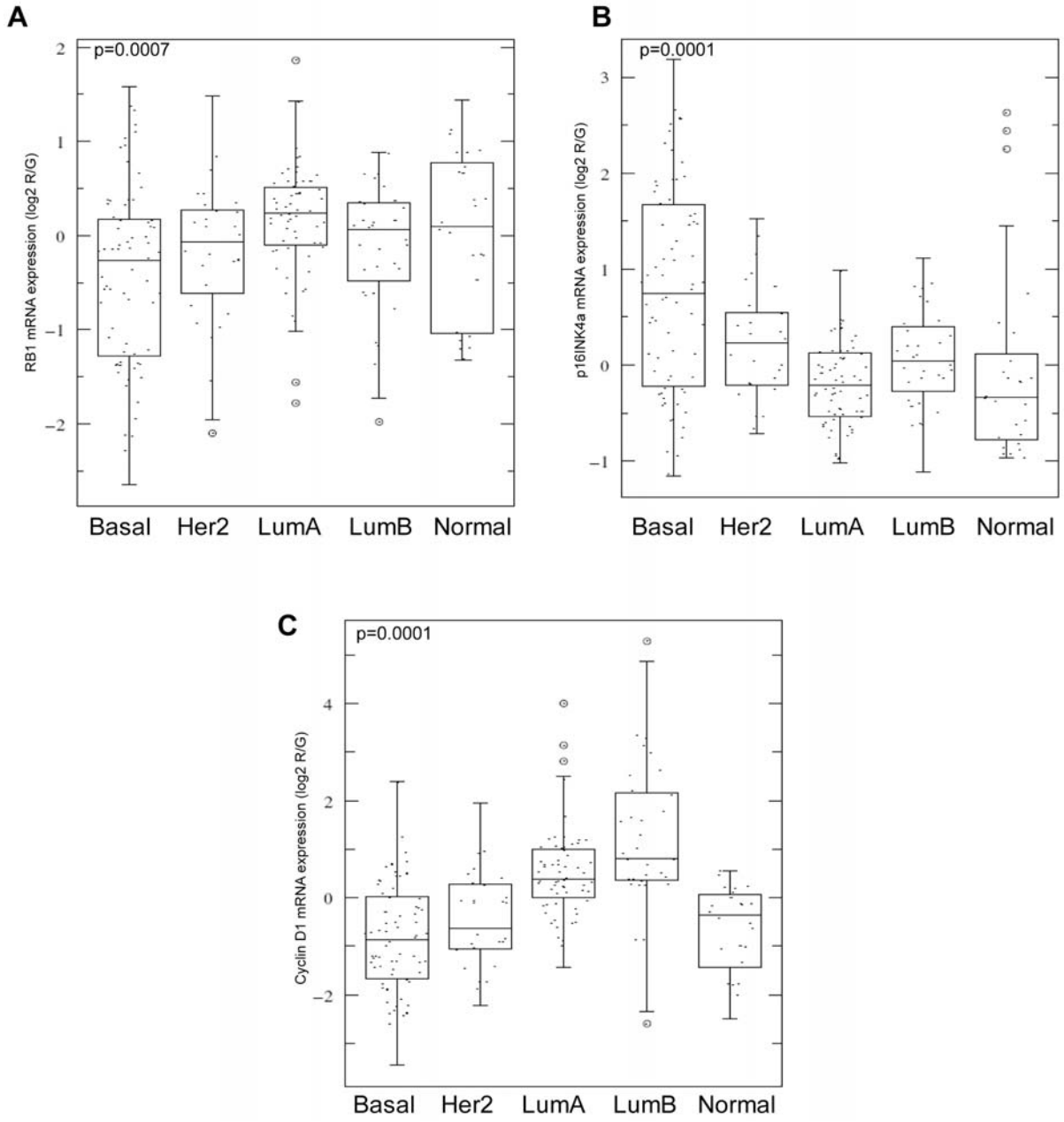

A
p=0.0007
RB1 mRNA expression (log2 R/G)
2
1
0
−1
−2
Basal
Her2
LumA
LumB
Normal
B
p=0.0001
p16INK4a mRNA expression (log2 R/G)
3
2
1
0
−1
Basal
Her2
LumA
LumB
Normal
C
p=0.0001
Cyclin D1 mRNA expression (log2 R/G)
4
2
0
−2
Basal
Her2
LumA
LumB
Normal

Figure 5.2 The expression of RB1, p16INK4a, and Cyclin D1 varies across the breast cancer intrinsic subtypes. Box plot comparisons of a) RB1, b) p16INK4a, and c) Cyclin D1 mRNA expression to five intrinsic subtypes as defined by 5-class SSP.

Next, using a previously defined proliferation gene signature [8], we clustered the gene expression data for this signature using just the 67 LOH informative patients (Figure 5.3A). This analysis was able to sort the samples into two groups with the right most group containing most of the RB1 LOH+ tumors (21/27 77.7%). The left most group contained 29/40 (72.5%) of the RB1 LOH-normal tumors. It is important to note, however, that many of the LOH+ tumors that did cluster in the left group, were located on the outer nodes and still showed high expression of the proliferation markers than the tumors in the center. Overall, ANOVA analysis showed that RB1 LOH is highly correlated with high expression of the proliferation gene cluster ($p=0.0001$) (Figure 5.3B).

LOH at the RB1 locus is associated with tumor subtype

We classified the tumors according to intrinsic subtype using the 5-class single sample predictor (SSP) from Hu et al[8]. The frequency of RB1 LOH varied by molecular subtype ($p=0.0002$) (Table 5.1). The lowest LOH frequencies were observed in the Luminal A (3/20 15%) and Normal-like (0/7 0%) subtypes, while the HER2+/ER- subtype had a frequency near the breast cancer average (3/9 33.3%). The highest frequency of RB1 LOH was observed in tumors of the basal-like subtype (13/18 72.2%) and the Luminal B subtype (8/13 61.5%), both of which are known to be highly proliferative tumor subtypes[8-10].

Figure 5.3

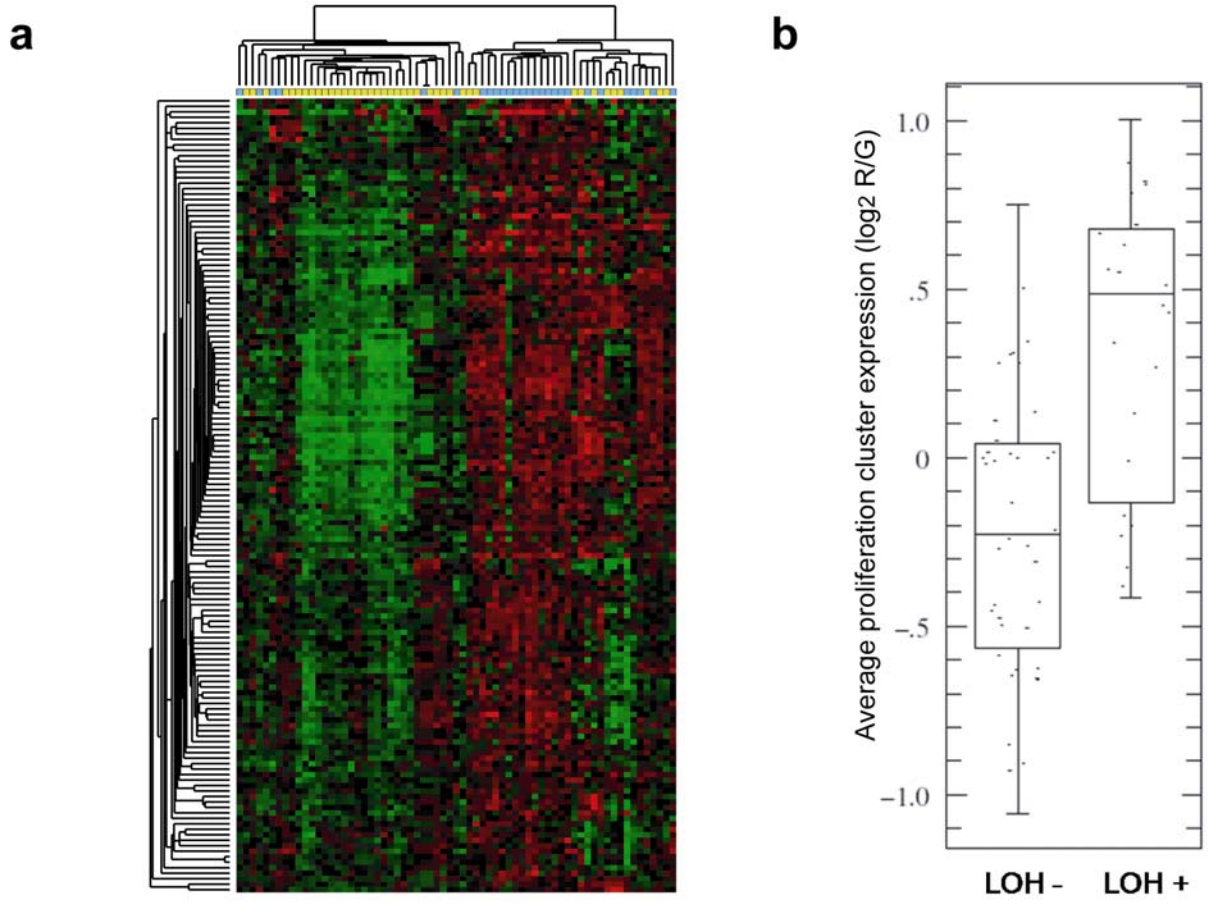

a
b
Average proliferation cluster expression (log2 R/G)
1.0
.5
0
-.5
-1.0
LOH -
LOH +

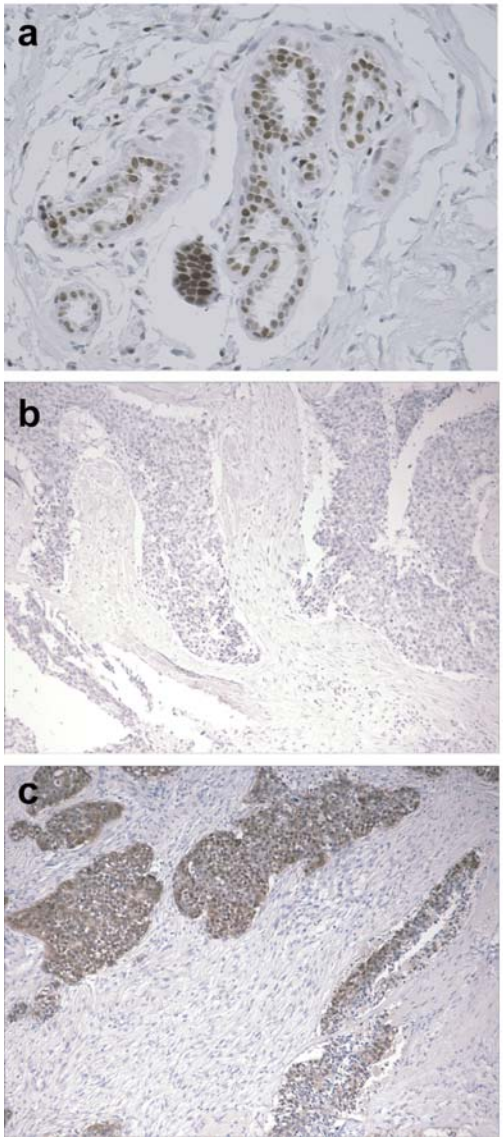
Figure 5.3 RB1 LOH is associated with high proliferation. a) two-way unsupervised hierchical clustering of breast tumor samples with informative RB1 LOH status (LOH + = blue, LOH - = yellow) using the proliferation gene cluster. b) Box plot comparison of the average proliferation cluster expression to RB1 LOH status ($p<0.0001$).

RB1 and P16INK4A immunostaining in breast carcinomas

RB1 is expressed ubiquitously in mammary epithelial cells and typically shows a nuclear staining pattern in normal human breast tissue (Figure 5.4a). RB1 immunostaining was statistically correlated with RB1 message levels (p=0.0081), however as has been described before, RB1 protein expression did not correlate with RB1 LOH (p=0.5) [5]; however, a trend for low expression of RB1 message to occur with RB1 LOH was observed (p=0.11). RB1 protein expression trended to be low in basal-like tumors (Figure 5.4b and Table 5.1), however, this relationship was also not statistically significant (p=0.064).

High p16INK4a staining (3+), which is likely a hallmark of lost RB1 function [16, 17], was seen in 32/119 tumors assayed and often included both nuclear and cytoplasmic staining. p16INK4a immunostaining was statistically correlated with p16INK4a message levels (p=0.013), especially when there was high staining observed (Appendix VA). p16INK4a immunostaining was also associated with molecular subtype (p=1.41E-05) with 22/33 (66.7%) of basal-like tumors showing 3+ staining (Figure 5.3c, Table 5.1). The correlation between RB1 LOH status and p16INK4a gene expression levels was also statistically significant (p=0.01) (Figure 5.3f). In addition, p16INK4a is similarly highly expressed in transgenic murine mammary tumors with loss of Rb function driven by SV40 large T-antigen or T121 (Appendix VB)[13].

Figure 5.4



d

Frequency Row Pct	P16-0	P16-1	P16-2	P16-3	Total
LOH-0	9 23.68	11 28.95	11 28.95	7 18.42	38
LOH-1	1 3.85	8 30.77	5 19.23	12 46.15	26
Total	10	19	16	19	64

Statistics for Table of response by subtype

Statistic	DF	Value	Prob
Chi-Square	3	8.4879	0.0369
Likelihood Ratio Chi-Square	3	9.2109	0.0266
Mantel-Haenszel Chi-Square	1	5.8308	0.0157
Phi Coefficient		0.3642	
Contingency Coefficient		0.3422	
Cramer's V		0.3642	

Fisher's Exact Test

Table Probability (P)	2.765E-04
Pr <= P	0.0363

Sample Size = 64

e

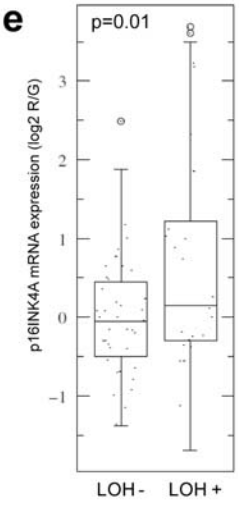


Figure 5.4 High p16INK4a mRNA and protein levels are associated with RB1 LOH. a) RB1 staining of normal breast tissue b) RB1 LOH + basal-like tumor lacking RB1 staining and c) showing heavy staining for p16INK4a both nuclear and cytoplasmic. d) comparison of p16INK4a IHC and RB1 LOH status. (p16 0=negative, 1=weak positive, 2=moderate positive, 3=strong positive) e) Box plot comparison showing high p16INK4a mRNA expression in RB1 LOH + breast tumors (p=0.01).

Table 5.1 Subtype specificity of RB1 LOH and P16INK4a IHC

Frequency Row Pct	Basal	Her2	LumA	LumB	Norm	Total
LOH+	13 48.15	3 11.11	3 11.11	8 29.63	0 0.00	27
LOH-	5 12.50	6 15.00	17 42.50	5 12.50	7 17.50	40
Total	18	9	20	13	7	67

Statistics for Table of response by subtype

Statistic	DF	Value	Prob
Chi-Square	4	20.2893	0.0004
Likelihood Ratio Chi-Square	4	23.3841	0.0001
Mantel-Haenszel Chi-Square	1	6.6034	0.0102
Phi Coefficient		0.5503	
Contingency Coefficient		0.4821	
Cramer's V		0.5503	

WARNING: 30% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Table Probability (P)	2.572E-07
Pr <= P	2.263E-04

Sample Size = 67

Frequency Row Pct	Basal	Her2	LumA	LumB	Norm	Total
RB0	6 42.86	2 14.29	3 21.43	3 21.43	0 0.00	14
RB1	13 29.55	5 11.36	14 31.82	4 9.09	8 18.18	44
RB2	5 12.82	6 15.38	14 35.90	6 15.38	8 20.51	39
RB3	7 35.00	2 10.00	4 20.00	7 35.00	0 0.00	20
Total	31	15	35	20	16	117

Statistics for Table of response by subtype

Statistic	DF	Value	Prob
Chi-Square	12	19.0589	0.0871
Likelihood Ratio Chi-Square	12	23.3889	0.0246
Mantel-Haenszel Chi-Square	1	0.9138	0.3391
Phi Coefficient		0.4036	
Contingency Coefficient		0.3743	
Cramer's V		0.2330	

WARNING: 40% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Table Probability (P)	1.881E-11
Pr <= P	0.0644

Sample Size = 117

Frequency Row Pct	Basal	Her2	LumA	LumB	Norm	Total
p16-0	2 8.00	4 16.00	8 32.00	6 24.00	5 20.00	25
p16-1	5 15.15	3 9.09	15 45.45	5 15.15	5 15.15	33
p16-2	4 13.79	4 13.79	9 31.03	6 20.69	6 20.69	29
p16-3	22 68.75	4 12.50	3 9.38	3 9.38	0 0.00	32
Total	33	15	35	20	16	119

Statistics for Table of response by subtype

Statistic	DF	Value	Prob
Chi-Square	12	43.2033	<.0001
Likelihood Ratio Chi-Square	12	45.0287	<.0001
Mantel-Haenszel Chi-Square	1	20.9443	<.0001
Phi Coefficient		0.6025	
Contingency Coefficient		0.5161	
Cramer's V		0.3479	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Table Probability (P)	1.407E-16
Pr <= P	1.410E-05

Sample Size = 119

RB1 LOH gene expression signature

To determine if there was a gene expression signature related to RB1 LOH, a 2-class Significance Analysis of Microarrays (SAM) was performed[18]. In total there were 452 genes that varied with RB1 LOH status with a false discovery rate of 0.94%, as compared to 11 genes identified at a FDR of 16.7% when a similar analysis was performed using RB immunostaining data (0 vs 1,2,and 3). 423 of these genes were highly expressed in tumors with RB1 LOH and an analysis of the Gene Ontologies associated with this gene set showed that cell cycle, cell division, DNA metabolism, spindle organization and biogenesis, and response to DNA damage were the top biological processes when using Bonferroni-corrected scores. Interestingly, E2F1, E2F3, and E2F5 are highly expressed in tumors with RB1 LOH. Also present in this list is RB1CC1, a regulator of RB1 expression that has been shown to contain truncating mutations in breast cancers[19]. We used whole genome RVista to calculate which transcription factor binding sites might be present within the 1000bp upstream regions of these genes [20] and determined that the top three transcription factor binding sites with $p<0.005$ were E2F4DP1, E2F1DP1RB, and E2F4DP2, showing that a majority of these genes are likely E2F-regulated; other statistically significant transcription factor binding sites were HIF1/ARHN-HIF1. Only 29 genes were significantly negatively/down regulated from the RB1 LOH SAM analysis, and there were no significant Gene Ontology categories enriched in this list.

RB1 LOH gene expression signature correlates with signatures of proliferation and RB-loss

Recently, an RB-loss gene expression signature was derived using mouse fibroblasts with either acute or chronic knockout of RB using RNAi[21]. As might be expected, this RB-loss signature was highly expressed in basal-like tumors (Appendix VD). This RB-loss signature showed overall in gene identify with the proliferation signature (29/139 genes of the RB-loss signature are contained in the 140 gene human proliferation signature used here), thus serving as further evidence that the proliferation signature contains many RB-E2F regulated genes. There was statistically significant overlap among all three RB-pathway signatures (RB-LOH, RB-loss, and proliferation) as determined by hypergeometric mean analysis $p<0.001$ (Figure 5.5, Appendix VC), thus all three signatures are likely tracking a common biology that is RB-E2F dependent. There were 20 genes which overlapped between all three gene lists, which included many cell cycle related genes including the spindle assembly checkpoint proteins BUB1 and MAD2 and many commonly used chemotherapeutic drug targets including TOP2A (doxorubicin, etoposide), thymidylate synthetase (5-FU), ribonucleotide reductase M2 (hydroxyurea), and CDC2 (flavopiridol, staurosporine).

Figure 5.5

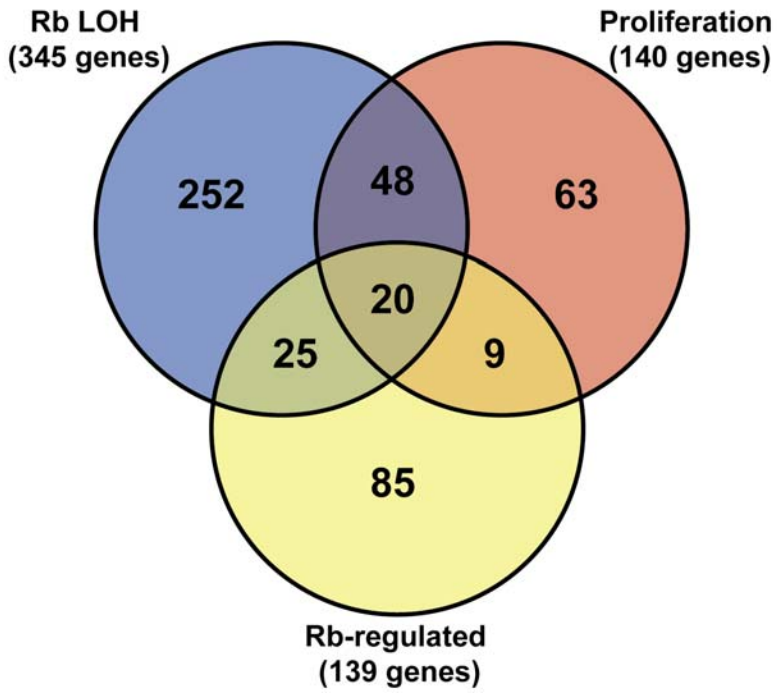


Figure 5.5 Venn diagram showing the significant overlap between the Rb LOH, proliferation, and Rb-regulated gene lists.

All four of these RB-pathway associated lists (RB-LOH, RB-loss, proliferation signature and 20 common genes) were highly predictive of breast cancer patient outcomes when using a 2-class, or 3-class, average value rank order expression cutoff and tested on the NKI295 patient data set (Figure 5.6 using RB-LOH list, all 4 lists give very similar results)[22], and on the Miller et al. 251 patient data set (data not shown)[23]. It should be noted, however, that multiple different investigators have independently identified different gene lists that contain a large number of so called proliferation/RB-pathway genes [12, 24-27], thus it is not unexpected that the RB-LOH signature is a strong prognostic profile.

Figure 5.6

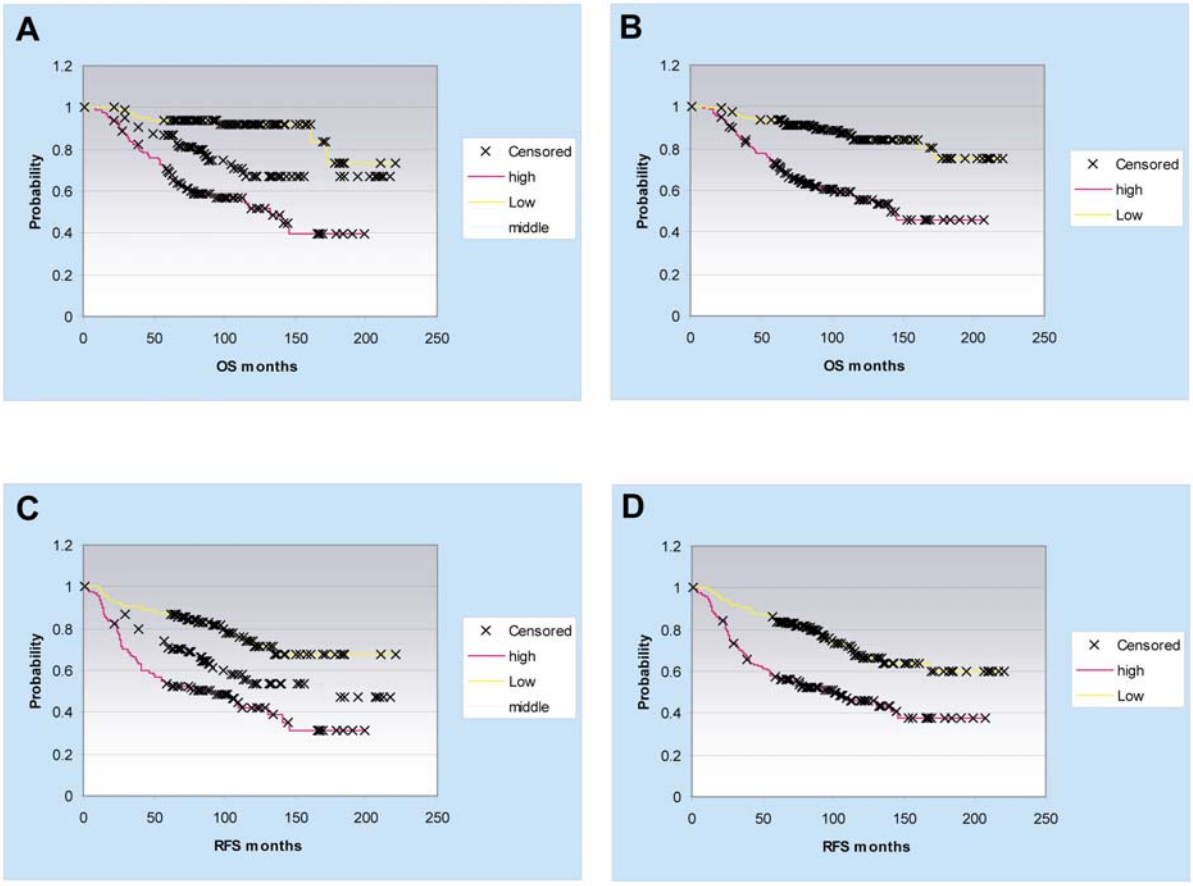


Figure 5.6 RB-LOH gene list is highly predictive of breast cancer patient outcome. Kaplan Meier survival curves looking at overall survival (OS) or relapse free survival (RFS) by dividing the patients in thirds (A,C) or in half (B,D) using the RB-LOH expression signature on the NKI295 breast cancer dataset

DISCUSSION

Our understanding of breast cancer biology has been improved by the identification of genomically defined tumor subtypes. These subtypes are defined by distinct gene expression patterns, molecular changes, and potential distinct developmental cell types of origin adding up to observed differences in outcome and responses to therapy. In this report, we show that the frequency of RB1 LOH observed varied significantly according to "intrinsic" subtype. RB1 LOH occurs at a frequency of 72.2% in basal-like breast tumors and 61.5% in Luminal B tumors, both of which are in the range of the 60-75% frequency observed in retinoblastomas[28-31]. RB1 protein staining as assessed by IHC, however, did not correlate with RB1 LOH in our study (as has been reported before); however, a SAM analysis supervised by RB1 LOH robustly identified many E2F genes, and E2F target genes, while the RB protein SAM analysis did not (data not shown), which suggests that RB1 LOH is a better biomarker of RB-pathway function than IHC staining for total RB protein.

Additional support for the functional loss of RB function in basal-like tumors comes from the correlation with high p16INK4a message and protein. The inverse relationship between p16INK4a and RB1 expression in breast cancers has been reported before[32, 33], however, this relationship and it association with basal-like tumors is new. Another intriguing link between p16INK4a and basal-like cells comes from studies on HMECs (human mammary epithelial cells), which have been shown to resemble the basal-like subtype by gene expression [34, 35]. It has been shown that in order for HMECs to proliferate in culture for an extended period *in vitro* they must overcome an RB-mediated

stress associated senescence barrier (stasis), which usually involves spontaneously losing p16 expression by promoter methylation [36]; the gene expression changes associated with this *in vitro* transition are similar to those we have reported occurring with RB1 LOH[37]. *In vivo* in basal-like tumors, however, the exact opposite appears to occur in that the RB-pathway barrier appears to be RB functional loss with a concomitant feedback loop that induces p16 gene and protein expression. The link between high p16 expression being caused by RB1 loss is known it has been shown that RB1 recruits Polycomb repression complexes to the p16 locus, which silence p16/INK4a transcription[38]. It is also well-known that cell cycle inhibition by p16INK4a is RB-dependent[39], and therefore, these RB1 deficient breast tumors would be expected to be refractory to the high amounts of p16INK4a present, which explains their high proliferation rates in the presence of the highest levels of p16. High p16INK4a expression has reproducibly been shown to be associated with poor prognosis [32, 40-42] and in a recent study by Grupka et al., p16INK4a staining of sentinel lymph nodes was predictive in determining the presence of non-sentinel node metastases[43]. In total, these data strongly argue that the RB-pathway lesion that occurs in most basal-like tumors is RB loss, with a compensatory activation of p16. Lastly, basal-like tumors also highly express a recently published RB-loss gene expression signature [21, 24], which we have shown to have significant similarity to the previously defined proliferation signature and our newly described list of genes that correlate with RB1 LOH. The elevated Ki-67 index seen in basal-like ductal carcinoma *in situ* lesions also suggests that RB1 loss may be an early event for this tumor type[44].

Similar to basal-like tumors, Luminal B tumors also showed a high frequency (61.5%) of RB1 LOH in our study, but did not show an association with high p16 expression. In the recent study by Bosco et al., luminal tumor derived cell lines were shown to be more proliferative and resistant to hormone therapy following knockdown of RB1[24], both of which are signatures of Luminal B tumors [8-10]. The RB-loss signature was shown to be predictive of outcome in a dataset containing only ER+ breast tumors treated with tamoxifen monotherapy. Therefore, the loss of RB1 function may also play a substantial role in the increased proliferation, possible resistance to hormonal therapies, and poor prognosis that is seen in Luminal B tumors. In addition, the knockdown of RB1 in established breast cancer cell lines has recently been shown to increase sensitivity to a variety of DNA-damaging therapeutic agents[24]. While these experiments were performed with ER+ tumor cell lines, it does open the possibility that the RB1 defect in basal-like tumors plays a role in their increased chemosensitivity compared to luminal tumors [45, 46].

The presence of LOH is typically thought to indicate that a mutated allele is present on the other chromosome and that the LOH unmasks the mutated allele. However, there is little evidence to suggest that dramatic structural changes aside from LOH are occurring at the RB1 locus in breast tumors. There are a few reports of alterations in RB1 in breast cancer with two reports showing structural changes as assessed by southern blotting in 7% and 19% of primary tumors[47, 48], and no published reports to our knowledge of point mutations. Interestingly, a study by Kallionemi et al. looking at RB1 loss in clinical breast cancer samples by fluorescent in situ hybridization showed that most of the cells within these tumors contained two copies of the RB1 gene even when they showed LOH by restriction fragment

length polymorphism at the RB1 locus[49]. When these studies are considered with the data presented here, they suggest a complex scenario where one allele is lost or altered in function by LOH, and the remaining allele/residual protein is compromised by yet to be identified mechanism(s) that potentially varies between tumor subtype, and potentially varies even within basal-like tumors; for example, some basal tumors with LOH show complete loss of RB protein, while others show high expression and both types show high proliferation. Thus additional studies will be needed to define these unconventional methods of RB1 inactivation.

In summary, here we show that RB1 LOH is a frequent occurrence in basal-like and luminal B breast tumors and is associated with deregulation of E2F-regulated genes and increased proliferation. Deregulation of the Rb pathway in cell lines has shown that it may be an important determinant of response to therapy[24]. Therefore, RB function would appear to be an important biomarker for informing treatment decisions. Future studies involving clinical trials are warranted to evaluate the impact RB1 loss has on therapeutic response in breast cancer overall and within molecular subtypes.

EXPERIMENTAL PROCEDURES

Patient samples and breast cancer microarray data sets

All human tumor samples were collected from fresh frozen primary breast tumors using IRB-approved protocols and were profiled as described previously[8, 13, 50, 51]. The clinical and pathological information for these samples can be obtained online shortly.

The primary microarray data for the 232 dataset is available at genome.unc.edu and in GEO under the accession number GSE3165. The data set containing only tumors with informative LOH status will be found in GEO shortly.

DNA isolation and detection of RB1 loss of heterozygosity

Patient DNA from lymphocytes and breast tumors was isolated using the DNeasy kit (Qiagen). We used two polymorphic markers, a variable number tandem repeat (VNTR) in intron 20 and D13S153; a microsatellite marker. The primers were previously published for intron 20 [52]. The primers for D13S153 (AFM058xd6a, AFM058xd6m) were obtained from the Genome Data Bank (http://www.gdb.org/) (Johns Hopkins University). The PCR products were run on the Agilent Bioanalyzer using DNA 1000 kit (Agilent). The patient was called informative when there were two alleles present in their normal DNA. LOH was called when there was a loss of an allele in the tumor for at least one of the two polymorphic sites.

Statistical Analysis

The Chi-Square test and Fisher-Freeman-Halton Exact test were used to examine correlations between RB1 LOH status, immunostaining, and tumor subtype using SAS 9.1 (Cary, NC). ANOVA analysis, unpaired Student's t-test and box plot were performed to compare RB1 LOH status or immunostaining, with gene expression using web based "Statistists to Use" at http://www.physics.csbsju.edu/stats/

SAM analysis was performed to identify genes that were significantly different between tumors with RB1 LOH + compared to LOH - tumors[18]. Expression analysis systemic explorer (EASE) was used to identify gene ontology categories overrepresented in the RB LOH gene list compared to the genes present on the array.

Whole genome RVista **(http://genome.lbl.gov/vista/index.shtml)** was used to determine the known transcription factor binding sites overrepresented in the 1kb upstream region in the lists of genes examined compared to the rest of the RefSeq genes in the whole human genome[20]

Hypergeometric mean analysis was performed as in Chung et al[53]. Briefly, the simulation was performed by randomly selecting a set of genes from the entire overlapping population of genes from which each pair of test gene sets were derived. The number randomly selected was set to the number of genes in our subset being tested. This randomly selected set was then compared for overlap, with the corresponding gene sets that were mentioned from published studies. The number of genes found in the overlap was recorded and the process was repeated 10,000 times. In the final step, the actual amount of overlap was compared between our gene set and the published gene sets, and this was compared to the simulated distribution. This comparison gives the likelihood of finding co-occurrences between these gene sets by chance. The simulation was performed independently for each pair of gene sets.

Immunohistochemistry

Formalin-fixed, paraffin-embedded tissue sections (~5um) were processed using standard immunostaining methods. Following deparaffinization in xylenes, slides were rehydrated through a graded series of alcohol and rinsed in phosphate buffered saline. Endogenous peroxidase activity was blocked with 3% hydrogen peroxidase. Samples were steamed for antigen retrieval with 10 mM citrate buffer (pH 6.0) for 30 min. Slides were then incubated for 20 minutes with diluted normal blocking serum. The sections were incubated for 60 minutes at room temperature with primary antibody pRb (Visionbiosystems Novocastra, NCL-L-RB-358 clone 13A10, 1:50 dilution) or p16 (Santa Cruz, H-156, 1:50 dilution). The slides were incubated for 45 minutes with diluted biotinylated secondary antibody (1:250 dilution) and 30 minutes with Vectastain Elite ABC reagent (Vector Laboratories). Sections were incubated in peroxidase substrate solution for visualization.. Slides were counterstained with hematoxylin and examined by light microscopy. Tumor immunoreactivity was scored 0=negative, 1=weak positive, 2=moderate positive, and 3=strong positive. The evaluation of p16 and Rb staining was performed by two independent researchers.

ACKNOWLEDGEMENTS

We thank Yue Xiong and Ned Sharpless for comments and reviews of this manuscript. C.M.P. was supported by funds from the NCI Breast SPORE program to UNC-CH (P50-CA58223-09A1), by RO1-CA-101227-01, the V Foundation for Cancer Research, and by the Breast Cancer Research Foundation.

REFERENCES

1. Sherr CJ: **Cancer cell cycles**. *Science* 1996, **274**(5293):1672-1677.

2. Santamaria D, Barriere C, Cerqueira A, Hunt S, Tardy C, Newton K, Caceres JF, Dubus P, Malumbres M, Barbacid M: **Cdk1 is sufficient to drive the mammalian cell cycle**. *Nature* 2007, **448**(7155):811-815.

3. Lee EY, To H, Shew JY, Bookstein R, Scully P, Lee WH: **Inactivation of the retinoblastoma susceptibility gene in human breast cancers**. *Science* 1988, **241**(4862):218-221.

4. Bieche I, Lidereau R: **Loss of heterozygosity at 13q14 correlates with RB1 gene underexpression in human breast cancer**. *Mol Carcinog* 2000, **29**(3):151-158.

5. Borg A, Zhang QX, Alm P, Olsson H, Sellberg G: **The retinoblastoma gene in breast cancer: allele loss is not correlated with loss of gene protein expression**. *Cancer Res* 1992, **52**(10):2991-2994.

6. Geradts J, Wilson PA: **High frequency of aberrant p16(INK4A) expression in human breast cancer**. *Am J Pathol* 1996, **149**(1):15-20.

7. Buckley MF, Sweeney KJ, Hamilton JA, Sini RL, Manning DL, Nicholson RI, deFazio A, Watts CK, Musgrove EA, Sutherland RL: **Expression and amplification of cyclin genes in human breast cancer**. *Oncogene* 1993, **8**(8):2127-2133.

8. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L *et al*: **The molecular portraits of breast tumors are conserved across microarray platforms**. *BMC Genomics* 2006, **7**(1):96.

9. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: **Repeated observation of breast tumor subtypes in independent gene expression data sets**. *Proc Natl Acad Sci U S A* 2003, **100**(14):8418-8423.

10. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc Natl Acad Sci U S A* 2001, **98**(19):10869-10874.

11. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, Perou CM: **Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma**. *Mod Pathol* 2005.

12. Whitfield ML, George LK, Grant GD, Perou CM: **Common markers of proliferation**. *Nat Rev Cancer* 2006, **6**(2):99-106.

13. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S *et al*: **Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors**. *Genome Biol* 2007, **8**(5):R76.

14. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T *et al*: **Genomic and transcriptional aberrations linked to breast cancer pathophysiologies**. *Cancer Cell* 2006, **10**(6):529-541.

15. Wang ZC, Lin M, Wei LJ, Li C, Miron A, Lodeiro G, Harris L, Ramaswamy S, Tanenbaum DM, Meyerson M *et al*: **Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers**. *Cancer Res* 2004, **64**(1):64-71.

16. Li Y, Nichols MA, Shay JW, Xiong Y: **Transcriptional repression of the D-type cyclin-dependent kinase inhibitor p16 by the retinoblastoma susceptibility gene product pRb**. *Cancer Res* 1994, **54**(23):6078-6082.

17. Tam SW, Shay JW, Pagano M: **Differential expression and cell cycle regulation of the cyclin-dependent kinase 4 inhibitor p16Ink4**. *Cancer Res* 1994, **54**(22):5816-5820.

18. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response**. *Proc Natl Acad Sci U S A* 2001, **98**(9):5116-5121.

19. Chano T, Kontani K, Teramoto K, Okabe H, Ikegawa S: **Truncating mutations of RB1CC1 in human breast cancer**. *Nat Genet* 2002, **31**(3):285-288.

20. Zambon AC, Zhang L, Minovitsky S, Kanter JR, Prabhakar S, Salomonis N, Vranizan K, Dubchak I, Conklin BR, Insel PA: **Gene expression patterns define key transcriptional events in cell-cycle regulation by cAMP and protein kinase A**. *Proc Natl Acad Sci U S A* 2005, **102**(24):8561-8566.

21. Markey MP, Bergseid J, Bosco EE, Stengel K, Xu H, Mayhew CN, Schwemberger SJ, Braden WA, Jiang Y, Babcock GF *et al*: **Loss of the retinoblastoma tumor suppressor: differential action on transcriptional programs related to cell cycle control and immune function**. *Oncogene* 2007.

22. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ *et al*: **A gene-expression signature as a predictor of survival in breast cancer**. *N Engl J Med* 2002, **347**(25):1999-2009.

23. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET *et al*: **An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival**. *Proc Natl Acad Sci U S A* 2005, **102**(38):13550-13555.

24. Bosco EE, Wang Y, Xu H, Zilfou JT, Knudsen KE, Aronow BJ, Lowe SW, Knudsen ES: **The retinoblastoma tumor suppressor modifies the therapeutic response of breast cancer**. *J Clin Invest* 2007, **117**(1):218-228.

25. Dai H, van 't Veer LJ, Lamb J, He YD, Mao M, Fine BM, Bernards R, van de Vijver MJ, Deutsch P, Sachs A *et al*: **A Cell Proliferation Signature is a Marker of Extremely Poor Outcome in a Subpopulation of Breast Cancer Patients**. *Proc Natl Acad Sci U S A* 2005, **65**:4059-4066.

26. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T *et al*: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer**. *N Engl J Med* 2004, **351**(27):2817-2826.

27. Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, Mone M, Hansen H, Buys SS, Rasmussen K *et al*: **Classification and risk stratification of**

invasive breast carcinomas using a real-time quantitative RT-PCR assay. *Breast Cancer Res* 2006, **8**(2):R23.

28. Choy KW, Pang CP, Yu CB, Wong HL, Ng JS, Fan DS, Lo KW, Chai JT, Wang J, Fu W *et al*: **Loss of heterozygosity and mutations are the major mechanisms of RB1 gene inactivation in Chinese with sporadic retinoblastoma**. *Hum Mutat* 2002, **20**(5):408.

29. Hagstrom SA, Dryja TP: **Mitotic recombination map of 13cen-13q14 derived from an investigation of loss of heterozygosity in retinoblastomas**. *Proc Natl Acad Sci U S A* 1999, **96**(6):2952-2957.

30. Ramprasad VL, Madhavan J, Murugan S, Sujatha J, Suresh S, Sharma T, Kumaramanickavel G: **Retinoblastoma in India: microsatellite analysis and its application in genetic counseling**. *Mol Diagn Ther* 2007, **11**(1):63-70.

31. Zhang XL, Fu WL, Zhao HX, Zhou LX, Huang JF, Wang JH: **Molecular studies of loss of heterozygosity in Chinese sporadic retinoblastoma patients**. *Clin Chim Acta* 2005, **358**(1-2):75-80.

32. Dublin EA, Patel NK, Gillett CE, Smith P, Peters G, Barnes DM: **Retinoblastoma and p16 proteins in mammary carcinoma: their relationship to cyclin D1 and histopathological parameters**. *Int J Cancer* 1998, **79**(1):71-75.

33. Gorgoulis VG, Koutroumbi EN, Kotsinas A, Zacharatos P, Markopoulos C, Giannikos L, Kyriakou V, Voulgaris Z, Gogas I, Kittas C: **Alterations of p16-pRb pathway and chromosome locus 9p21-22 in sporadic invasive breast carcinomas**. *Mol Med* 1998, **4**(12):807-822.

34. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**(6797):747-752.

35. Ross DT, Perou CM: **A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines**. *Dis Markers* 2001, **17**(2):99-109.

36. Brenner AJ, Stampfer MR, Aldaz CM: **Increased p16 expression with first senescence arrest in human mammary epithelial cells and extended growth capacity with p16 inactivation**. *Oncogene* 1998, **17**(2):199-205.

37. Li Y, Pan J, Li JL, Lee JH, Tunkey C, Saraf K, Garbe JC, Whitley MZ, Jelinsky SA, Stampfer MR *et al*: **Transcriptional changes associated with breast cancer occur as normal human mammary epithelial cells overcome senescence barriers and become immortalized**. *Mol Cancer* 2007, **6**:7.

38. Kotake Y, Cao R, Viatour P, Sage J, Zhang Y, Xiong Y: **pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16INK4alpha tumor suppressor gene**. *Genes Dev* 2007, **21**(1):49-54.

39. Lukas J, Parry D, Aagaard L, Mann DJ, Bartkova J, Strauss M, Peters G, Bartek J: **Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16**. *Nature* 1995, **375**(6531):503-506.

40. Emig R, Magener A, Ehemann V, Meyer A, Stilgenbauer F, Volkmann M, Wallwiener D, Sinn HP: **Aberrant cytoplasmic expression of the p16 protein in breast cancer is associated with accelerated tumour proliferation**. *Br J Cancer* 1998, **78**(12):1661-1668.

41. Hui R, Macmillan RD, Kenny FS, Musgrove EA, Blamey RW, Nicholson RI, Robertson JF, Sutherland RL: **INK4a gene expression and methylation in primary breast cancer: overexpression of p16INK4a messenger RNA is a marker of poor prognosis**. *Clin Cancer Res* 2000, **6**(7):2777-2787.

42. Milde-Langosch K, Bamberger AM, Rieck G, Kelp B, Loning T: **Overexpression of the p16 cell cycle inhibitor in breast cancer is associated with a more malignant phenotype**. *Breast Cancer Res Treat* 2001, **67**(1):61-70.

43. Grupka NL, Bloom C, Singh M: **Expression of retinoblastoma protein in breast cancer metastases to sentinel nodes: evaluation of its role as a marker for the presence of metastases in non-sentinel axillary nodes, and comparison to p16INK4a**. *Appl Immunohistochem Mol Morphol* 2006, **14**(1):63-70.

44. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, Tse CK, Nyante S, Millikan RC: **Identification of a basal-like subtype of breast ductal carcinoma in situ**. *Hum Pathol* 2007, **38**(2):197-204.

45. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM: **The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes**. *Clin Cancer Res* 2007, **13**(8):2329-2334.

46. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P *et al*: **Breast cancer molecular subtypes respond differently to preoperative chemotherapy**. *Clin Cancer Res* 2005, **11**(16):5678-5685.

47. T'Ang A, Varley JM, Chakraborty S, Murphree AL, Fung YK: **Structural rearrangement of the retinoblastoma gene in human breast carcinoma**. *Science* 1988, **242**(4876):263-266.

48. Varley JM, Armour J, Swallow JE, Jeffreys AJ, Ponder BA, T'Ang A, Fung YK, Brammar WJ, Walker RA: **The retinoblastoma gene is frequently altered leading to loss of expression in primary breast tumours**. *Oncogene* 1989, **4**(6):725-729.

49. Kallioniemi A, Kallioniemi OP, Waldman FM, Chen LC, Yu LC, Fung YK, Smith HS, Pinkel D, Gray JW: **Detection of retinoblastoma gene copy number in metaphase chromosomes and interphase nuclei by fluorescence in situ hybridization**. *Cytogenet Cell Genet* 1992, **60**(3-4):190-193.

50. Oh DS, Troester MA, Usary J, Hu Z, He X, Fan C, Wu J, Carey LA, Perou CM: **Estrogen-Regulated Genes Predict Survival in Hormone Receptor-Positive Breast Cancers**. *J Clin Oncol* 2006.

51. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, Glas AM, Perou CM, Van't Veer LJ: **Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer**. *Cancer Res* 2005, **65**(20):9155-9158.

52. Semczuk A, Marzec B, Roessner A, Jakowicki JA, Wojcierowski J, Schneider-Stock R: **Loss of heterozygosity of the retinoblastoma gene is correlated with the**

altered pRb expression in human endometrial cancer. *Virchows Arch* 2002, **441**(6):577-583.

53. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X *et al*: **Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression**. *Cancer Cell* 2004, **5**(5):489-500.

CHAPTER VI

CONCLUSIONS

There are many model systems in which to study breast cancer, each with their inherent advantages and disadvantages[1]. We turned to the mouse because we desired a model that will give us the ability to study spontaneous tumor development and therapy response *in vivo*. Furthermore, the mouse is amenable to genetic manipulation. Due to the heterogeneity of breast cancer, one model is not adequate. Luckily, there are many mouse mammary tumor models that have been generated over the past twenty years. In order to use these models in the most relevant way, the question we first had to answer was how do these mouse models relate to the human subtypes of breast cancer? Which models mimic aspects of human breast cancer and therefore may be useful to further study in a relevant way? It is not always clear which features of a human cancer are most relevant for disease comparisons (e.g. genetic aberrations, histological features, tumor biology). Genomic profiling provided us with a tool for comparative cancer analysis and offered a powerful means of cross-species comparison. We used gene expression analysis to classify a large set of mouse mammary tumor models and human breast tumors. The results provided us with biological insights among and across the mouse models, and comparisons with human data have identified biologically and clinically significant shared features.

These studies have highlighted what we feel are the two major applications for genomic comparisons between human tumors and their potential murine counterparts. First, these studies have identified those models that contain individual and/or global characteristics of particular classes of human tumors. Examples of important across-species biological characteristics identified in our studies included the classification of murine tumors, like human tumors, into basal and luminal groups. In our original study, four murine models developed potential luminal-like tumors (TgMMTV-*Neu*, TgMMTV-*PyMT*, TgWAP-*Myc*, and TgWAP-*Int3*), which is not surprising since both MMTV and WAP are thought to direct expression in differentiated alveolar/luminal cells [2, 3]. The luminal profile in the mouse models, however, lacked the expression of ER and ER-regulated genes and thus, there appears to be no murine counterpart of human Luminal A tumors.

Several murine models did show expression features consistent with human basal-like tumors including the TgC3(1)-*Tag*, TgWAP-*Tag* and *Brca1*-deficient models. The SV40 T-antigen used in the TgC3(1)-*Tag* and TgWAP-*Tag* models inactivates p53 and RB, which also appear to be two likely events that occur in human basal-like tumors because these tumors are known to harbor *p53* mutations[4], have high mitotic grade and the highest expression of proliferation genes [4, 5], which are known E2F targets[6]. The proliferation signature in human breast cancers is itself prognostic[7], and is also predictive of benefit to chemotherapy in ER+ patients [8].

We found that *Brca1* loss (coincident with *p53* mutation), in mice gives rise to tumors with features of basal-like human tumors. This finding was notable because human *BRCA1* germline mutation carriers almost exclusively develop basal-like tumors[5, 9]. In addition, most human *BRCA1* mutant tumors are p53-deficient[10, 11], which is required in these models to generate tumors with a reasonable latency and penetrance. Interestingly, a recent report also shows that conditional loss of BRCA1 and p53 using K14-Cre in mice induces mammary tumors also with a human basal-like breast cancer phenotype[12]. This looks to be a more consistent phenotype than the MMTV-Cre model showing the importance of the choice of promoter.

These data suggest a conserved predisposition of the basal-like cell type, or its antecedent cell, to transform as a result of *BRCA1*, *TP53*, and *RB*-pathway loss. Most DMBA-induced carcinomas also showed basal-like cell lineage features, suggesting that this cell type is also susceptible to DMBA-mediated tumorigenesis. Finally, some TgMMTV-*Wnt1* tumors showed a combination of basal-like and luminal characteristics by gene expression, which is consistent with the observation that tumors of this model generally contain cells from both mammary epithelial lineages[13].

With these luminal and basal distinctions, our comparison of mouse models has begun to hint at the cell types of origin of these tumors and the human tumors they mimic. Studying the cellular origins of cancers, whether the result of transformation of a stem cell,

progenitor cell, or de-differentiation of a mature cell, is a challenging task in human model systems and mouse models should be more amenable to this task.

The second major result of this comparative study was showing that analyses of murine models can inform the human disease and guide further discovery. An example of murine models informing the human disease is encompassed by the analysis of the new potential human subtype discovered here which we refer to as claudin-low. We have begun to further analyze these tumors to confirm that this is a *bona fide* subtype. The statistically significant gene overlap with a histologically distinct subset of murine tumors suggests that it is a distinct biological entity. This human subtype is defined by the low expression of genes involved in tight junctions and cell-cell adhesion including Claudins 3 and 12, Occludin, and E-cadherin. These tumors also show a marked increased expression of lymphocyte markers validated by the increased tumor infiltrating lymphocytes seen on examination. Consistent with their transcriptional profiles, immunohistochemical analysis showed that most claudin-low tumors expressed Claudin 3 and E-cadherin at low levels.

A second example of the murine tumors guiding discovery in humans was the common association of a K-Ras containing amplicon in a subset of human basal-like tumors and in the murine basal-like TgC3(1)-*Tag* strain tumors. Previous reports have shown that K-Ras is important for tumor progression in the TgC3(1)-*Tag* model. We have now begun to look further for copy number gains and losses in regions of conserved synteny that are in common between tumors arising in mouse models and human breast tumor subtypes. We

believe this will lead to the ability to narrow down the size of these regions and ultimately pinpoint the important oncogenes and tumor suppressors. These models will then serve as *in vivo* experimental systems to explore mechanistically.

A third example of the mouse guiding discovery was the high expression of the proliferation cluster seen in common between murine models initiated using T-antigen or T121 and human basal-like tumors. This suggested to us that basal-like tumors may also have a defect in the RB-E2F pathway. We evaluated the RB1 locus and observed that RB1 LOH varies significantly across human breast tumor subtype. RB1 LOH occurs at a frequency of 72.2% in basal-like breast tumors and 61.5% in Luminal B tumors, both of which are in the range of the 60-75% frequency observed in human retinoblastomas[14-17].

Knockdown of RB1 in established breast cancer cell lines has recently been shown to increase sensitivity to a variety of DNA-damaging therapeutic agents[18]. While these experiments were performed with ER+ tumor cell lines, it does open the possibility that the RB1 defect in basal-like tumors plays a role in their increased chemosensitivity compared to luminal tumors. In recent studies, basal-like tumors have been shown to have a higher sensitivity to neoadjuvant chemotherapy, however patients without complete response have a poor outcome[19, 20]. Some possibilities among many for the incomplete responders doing poorly could reflect the fact that RB1 loss and checkpoint loss can also lead to an increase in mutation rate in surviving cells or these tumors could have intact RB1. Thus a remaining

important question is whether RB1 functional status is involved in sensitivity to chemotherapy within the basal-like subtype.

Similar to basal-like tumors, Luminal B tumors showed a high frequency (61.5%) of RB1 LOH in our study. In the recent study by Bosco et al., luminal tumor cell lines were shown to be more proliferative and resistant to hormone therapy following knockdown of RB1[18]. The RB loss signature was shown to be predictive of outcome in a dataset containing only ER+ breast tumors treated with tamoxifen monotherapy. Therefore, the loss of RB1 may also play a substantial role in the increased proliferation, possible resistance to hormonal therapies, and poor prognosis that is seen in Luminal B tumors as compared to Luminal A tumors.

An important caveat to all comparative studies is that there are clear biological differences between mice and humans, which may or may not directly impact disease mechanisms. A potential example of inherent species difference could be the aforementioned biology associated with ER and its downstream pathway. In humans, ER is highly expressed in luminal tumors[21], with the luminal phenotype being characterized by the high expression of some genes that are ER-regulated like *PR* and *RERG[22]*, and other luminal genes that are likely GATA3-regulated including *AGR2* and *K8/18*[23]. In our original mouse model study, ER expression was low to absent in all the tumors we tested, as was the expression of most human ER-responsive genes. This finding is consistent with previous reports that most late-stage murine mammary tumors are ER-negative ([24] and references

within). However, it should be noted that two human luminal tumor-defining genes (*XBP1* and *GATA3*[23], were both highly expressed in murine luminal tumors. Taken together, these data suggest that the human "luminal" profile may actually be a combination of at least two profiles, one of which is ER-regulated and another of which is GATA3-regulated. Support for a link between *GATA3* and luminal cell origins comes from studies of *GATA3* loss in mice. The selective loss of *GATA3* in the mammary gland resulted in either a lack of luminal cells, or a significant decrease in the number and/or maturation of luminal cells[25, 26]. In a recent study it was shown that *GATA3*'s transcriptional regulation of the ER-alpha gene itself is required for estradiol stimulation of cell cycle progression[27]. Reciprocally, ER-alpha directly stimulates the transcription of *GATA3*. Why and how these mouse tumor models express *GATA3* without ER is an open question for exploration? These mouse models may also represent a unique opportunity to explore *GATA3* function independent of ER. These results suggest that in the mouse models tested, the ER-regulated gene cassette that is present in human luminal tumors is missing, and that the GATA3-mediated luminal signature remains. We believe, due to the partial human luminal tumor signature seen in mice, that the murine luminal models, especially the TgMMTV-*Neu* model, best resemble human luminal tumors and more specifically luminal B tumors, which are luminal tumors that express low amounts of ER and show a poor outcome[4, 5, 28].

A few murine ER-positive mammary tumor models have been developed[29-32]. We have analyzed a number of these models and they do not seem to resemble ER+ luminal human tumors. Therefore, the mouse may not be a faithful model in which to study estrogen-

dependent breast cancer and one may have to turn to the rat as a model. Many of these models constitute a new class of mouse mammary tumors that have alveolar expression features, a subtype that does not appear to have a common human correlate.

In a recent population-based epidemiological study, pregnancy was shown to be protective for human luminal A tumors with increased parity and younger age at first full-term pregnancy each showing a reduced risk[33]. On the other hand, pregnancy was shown to be associated with basal-like breast cancer with parity and younger age at first full-term pregnancy increasing risk. In another study, increasing age at menarche was associated with a reduced risk of basal-like breast cancer[34]. Therefore, it appears that hormone exposure is protective against luminal breast cancer but constitutes a risk factor for basal-like breast cancer. Recently, exogenous estrogen and progesterone exposure for short durations was shown to be able to mimic the protective effects of pregnancy in two genetically engineered mammary tumor models[35]. One of the models used was MMTV-LTR driving the activated Neu transgene, similar to the Neu model that we showed had luminal expression features in our study. The other model was the p53 transplant model. In a separate study, targeted conditional ER-alpha overexpression in mammary glands of TgMMTV-Cre;Brca1 p53, the same mice we showed often have basal features, increased the incidence and number of mammary tumors[36]. Many of these tumors were ER-negative even though ER was inducibly overexpressed in this model. Straightforward parity studies cannot be performed in many of the mouse models in current use because the MMTV promoter (and WAP) is

induced during pregnancy. However, the results of these studies are promising for modeling hormone and other environmental exposures in the mouse with relevance to human subtypes.

One of the key aspects of credentialing a model is to determine if it behaves like its human counterpart. This is particularly important when considering the use of mouse models preclinically. There have been a number of recent success stories of GEMs showing drug efficacy including mice with mutant EGFR-induced lung adenocarcinomas responding favorably to EGFR-targeted therapies [37, 38]. Our across species comparison showed us that the TgMMTV-*Neu* model exhibits gene expression patterns similar to human luminal tumors. A recent report showed that treatment of this same mouse model with the mTOR inhibitor, rapamycin, led to growth arrest and regression of the primary tumors while inhibiting proliferation in lung metastases[39]. Taken together, these studies suggest that human luminal tumors and especially the subset with HER2 overexpression or PIK3CA or PTEN mutations may be particularly sensitive to mTOR inhibitors.

The analyses described here identified many common and biologically relevant signatures across tumors of both species. These analyses also confirm the notion that there is not a single murine model that can faithfully recapitulate all of the complexities of human breast cancer, or even represent a human breast cancer subtype, however, the murine models do show shared features with specific human subtypes. It is these commonalties that have provided insights into important tumor biology and laid the groundwork for many future studies.

REFERENCES

1. Vargo-Gogola T, Rosen JM: **Modelling breast cancer: one size does not fit all**. *Nat Rev Cancer* 2007, **7**(9):659-672.

2. Hennighausen LG, Sippel AE: **Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins**. *Nucleic Acids Res* 1982, **10**(8):2677-2684.

3. Munoz B, Bolander FF, Jr.: **Prolactin regulation of mouse mammary tumor virus (MMTV) expression in normal mouse mammary epithelium**. *Mol Cell Endocrinol* 1989, **62**(1):23-29.

4. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc Natl Acad Sci U S A* 2001, **98**(19):10869-10874.

5. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: **Repeated observation of breast tumor subtypes in independent gene expression data sets**. *Proc Natl Acad Sci U S A* 2003, **100**(14):8418-8423.

6. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Barrette TR, Ghosh D, Chinnaiyan AM: **Mining for regulatory programs in the cancer transcriptome**. *Nat Genet* 2005, **37**(6):579-583.

7. Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, Mone M, Hansen H, Buys SS, Rasmussen K *et al*: **Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay**. *Breast Cancer Res* 2006, **8**(2):R23.

8. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J *et al*: **Gene Expression and Benefit of Chemotherapy in Women With Node-Negative, Estrogen Receptor-Positive Breast Cancer**. *J Clin Oncol* 2006.

9. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, Trudel M, Akslen LA: **Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer**. *J Natl Cancer Inst* 2003, **95**(19):1482-1485.

10. Crook T, Brooks LA, Crossland S, Osin P, Barker KT, Waller J, Philp E, Smith PD, Yulug I, Peto J *et al*: **p53 mutation with frequent novel condons but not a mutator phenotype in BRCA1- and BRCA2-associated breast tumours**. *Oncogene* 1998, **17**(13):1681-1689.

11. Phillips KA, Nichol K, Ozcelik H, Knight J, Done SJ, Goodwin PJ, Andrulis IL: **Frequency of p53 mutations in breast carcinomas from Ashkenazi Jewish carriers of BRCA1 mutations**. *J Natl Cancer Inst* 1999, **91**(5):469-473.

12. Liu X, Holstege H, van der Gulden H, Treur-Mulder M, Zevenhoven J, Velds A, Kerkhoven RM, van Vliet MH, Wessels LF, Peterse JL *et al*: **Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer**. *Proc Natl Acad Sci U S A* 2007, **104**(29):12111-12116.

13. Li Y, Welm B, Podsypanina K, Huang S, Chamorro M, Zhang X, Rowlands T, Egeblad M, Cowin P, Werb Z *et al*: **Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells**. *Proc Natl Acad Sci U S A* 2003, **100**(26):15853-15858.

14. Choy KW, Pang CP, Yu CB, Wong HL, Ng JS, Fan DS, Lo KW, Chai JT, Wang J, Fu W *et al*: **Loss of heterozygosity and mutations are the major mechanisms of RB1 gene inactivation in Chinese with sporadic retinoblastoma**. *Hum Mutat* 2002, **20**(5):408.

15. Hagstrom SA, Dryja TP: **Mitotic recombination map of 13cen-13q14 derived from an investigation of loss of heterozygosity in retinoblastomas**. *Proc Natl Acad Sci U S A* 1999, **96**(6):2952-2957.

16. Ramprasad VL, Madhavan J, Murugan S, Sujatha J, Suresh S, Sharma T, Kumaramanickavel G: **Retinoblastoma in India: microsatellite analysis and its application in genetic counseling**. *Mol Diagn Ther* 2007, **11**(1):63-70.

17. Zhang XL, Fu WL, Zhao HX, Zhou LX, Huang JF, Wang JH: **Molecular studies of loss of heterozygosity in Chinese sporadic retinoblastoma patients**. *Clin Chim Acta* 2005, **358**(1-2):75-80.

18. Bosco EE, Wang Y, Xu H, Zilfou JT, Knudsen KE, Aronow BJ, Lowe SW, Knudsen ES: **The retinoblastoma tumor suppressor modifies the therapeutic response of breast cancer**. *J Clin Invest* 2007, **117**(1):218-228.

19. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM: **The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes**. *Clin Cancer Res* 2007, **13**(8):2329-2334.

20. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P *et al*: **Breast cancer molecular subtypes respond differently to preoperative chemotherapy**. *Clin Cancer Res* 2005, **11**(16):5678-5685.

21. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**(6797):747-752.

22. Oh DS, Troester MA, Usary J, Hu Z, He X, Fan C, Wu J, Carey LA, Perou CM: **Estrogen-Regulated Genes Predict Survival in Hormone Receptor-Positive Breast Cancers**. *J Clin Oncol* 2006.

23. Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, Langerod A, Karesen R, Oh DS, Dressler LG *et al*: **Mutation of GATA3 in human breast tumors**. *Oncogene* 2004, **23**(46):7669-7678.

24. Medina D: **Mammary developmental fate and breast cancer risk**. *Endocr Relat Cancer* 2005, **12**(3):483-495.

25. Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z: **GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland**. *Cell* 2006, **127**(5):1041-1055.

26. Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, Hartley L, Robb L, Grosveld FG, van der Wees J *et al*: **Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation**. *Nat Cell Biol* 2006.

27. Eeckhoute J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M: **Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer**. *Cancer Res* 2007, **67**(13):6477-6483.

28. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L *et al*: **The molecular portraits of breast tumors are conserved across microarray platforms**. *BMC Genomics* 2006, **7**(1):96.

29. Wijnhoven SW, Zwart E, Speksnijder EN, Beems RB, Olive KP, Tuveson DA, Jonkers J, Schaap MM, van den Berg J, Jacks T *et al*: **Mice expressing a mammary gland-specific R270H mutation in the p53 tumor suppressor gene mimic human breast cancer development**. *Cancer Res* 2005, **65**(18):8166-8173.

30. Frech MS, Halama ED, Tilli MT, Singh B, Gunther EJ, Chodosh LA, Flaws JA, Furth PA: **Deregulated estrogen receptor alpha expression in mammary epithelial cells of transgenic mice results in the development of ductal carcinoma in situ**. *Cancer Res* 2005, **65**(3):681-685.

31. Lin SC, Lee KF, Nikitin AY, Hilsenbeck SG, Cardiff RD, Li A, Kang KW, Frank SA, Lee WH, Lee EY: **Somatic mutation of p53 leads to estrogen receptor alpha-positive and -negative mouse mammary tumors with high frequency of metastasis**. *Cancer Res* 2004, **64**(10):3525-3532.

32. Torres-Arzayus MI, De Mora JF, Yuan J, Vazquez F, Bronson R, Rue M, Sellers WR, Brown M: **High tumor incidence and activation of the PI3K/AKT pathway in transgenic mice define AIB1 as an oncogene**. *Cancer Cell* 2004, **6**(3):263-274.

33. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Smith LV, Labbok MH, Geradts J, Bensen JT, Jackson S *et al*: **Epidemiology of basal-like breast cancer**. *Breast Cancer Res Treat* 2007.

34. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A *et al*: **Differences in risk factors for breast cancer molecular subtypes in a population-based study**. *Cancer Epidemiol Biomarkers Prev* 2007, **16**(3):439-443.

35. Rajkumar L, Kittrell FS, Guzman RC, Brown PH, Nandi S, Medina D: **Hormone-induced protection of mammary tumorigenesis in genetically engineered mouse models**. *Breast Cancer Res* 2007, **9**(1):R12.

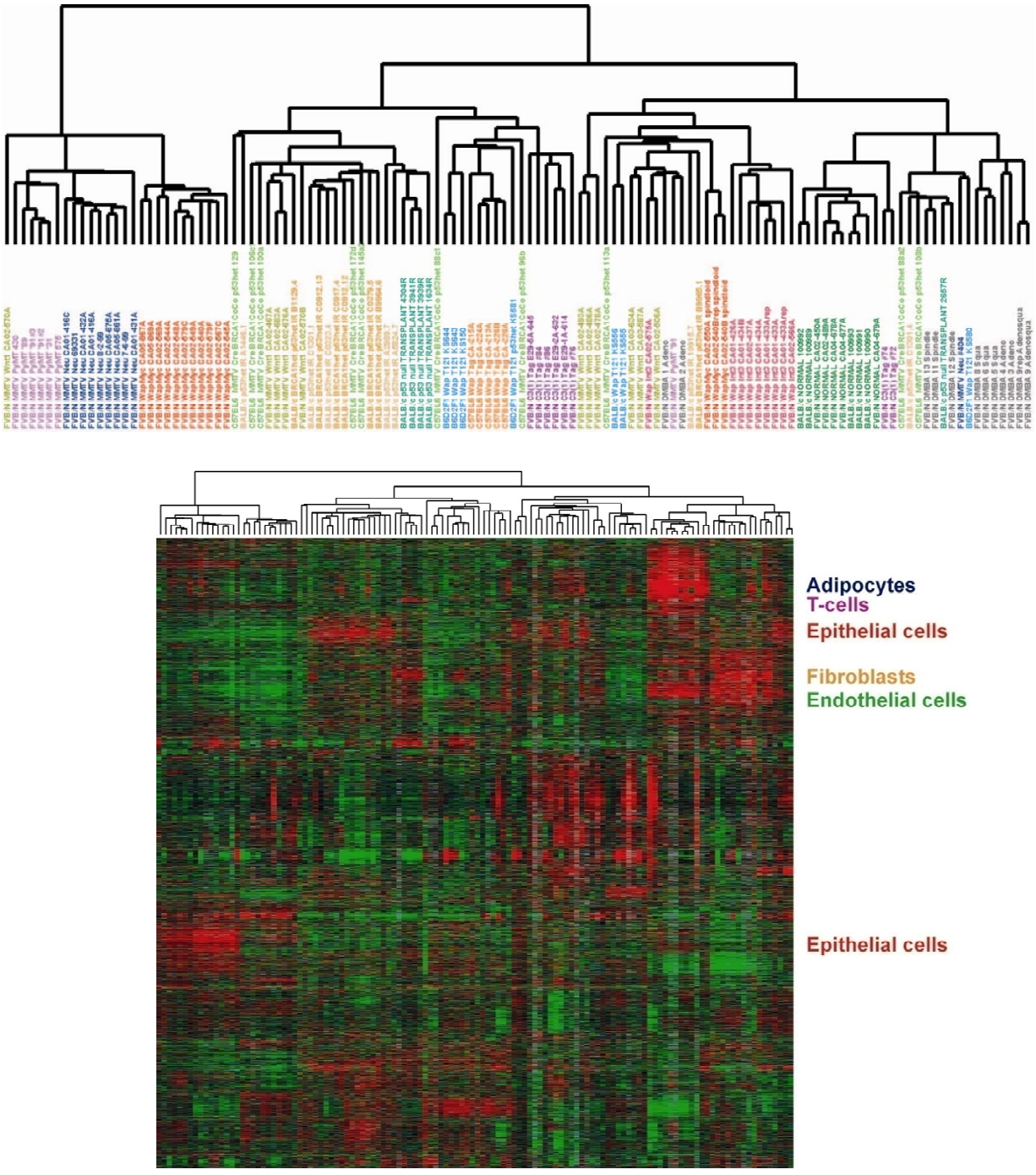
36. Jones LP, Tilli MT, Assefnia S, Torre K, Halama ED, Parrish A, Rosen EM, Furth PA: **Activation of estrogen signaling pathways collaborates with loss of Brca1 to promote development of ERalpha-negative and ERalpha-positive mammary preneoplasia and cancer**. *Oncogene* 2007.

37. Politi K, Zakowski MF, Fan PD, Schonfeld EA, Pao W, Varmus HE: **Lung adenocarcinomas induced in mice by mutant EGF receptors found in human lung cancers respond to a tyrosine kinase inhibitor or to down-regulation of the receptors**. *Genes Dev* 2006, **20**(11):1496-1510.

38. Ji H, Li D, Chen L, Shimamura T, Kobayashi S, McNamara K, Mahmood U, Mitchell A, Sun Y, Al-Hashem R *et al*: **The impact of human EGFR kinase domain mutations on lung tumorigenesis and in vivo sensitivity to EGFR-targeted therapies**. *Cancer Cell* 2006, **9**(6):485-495.

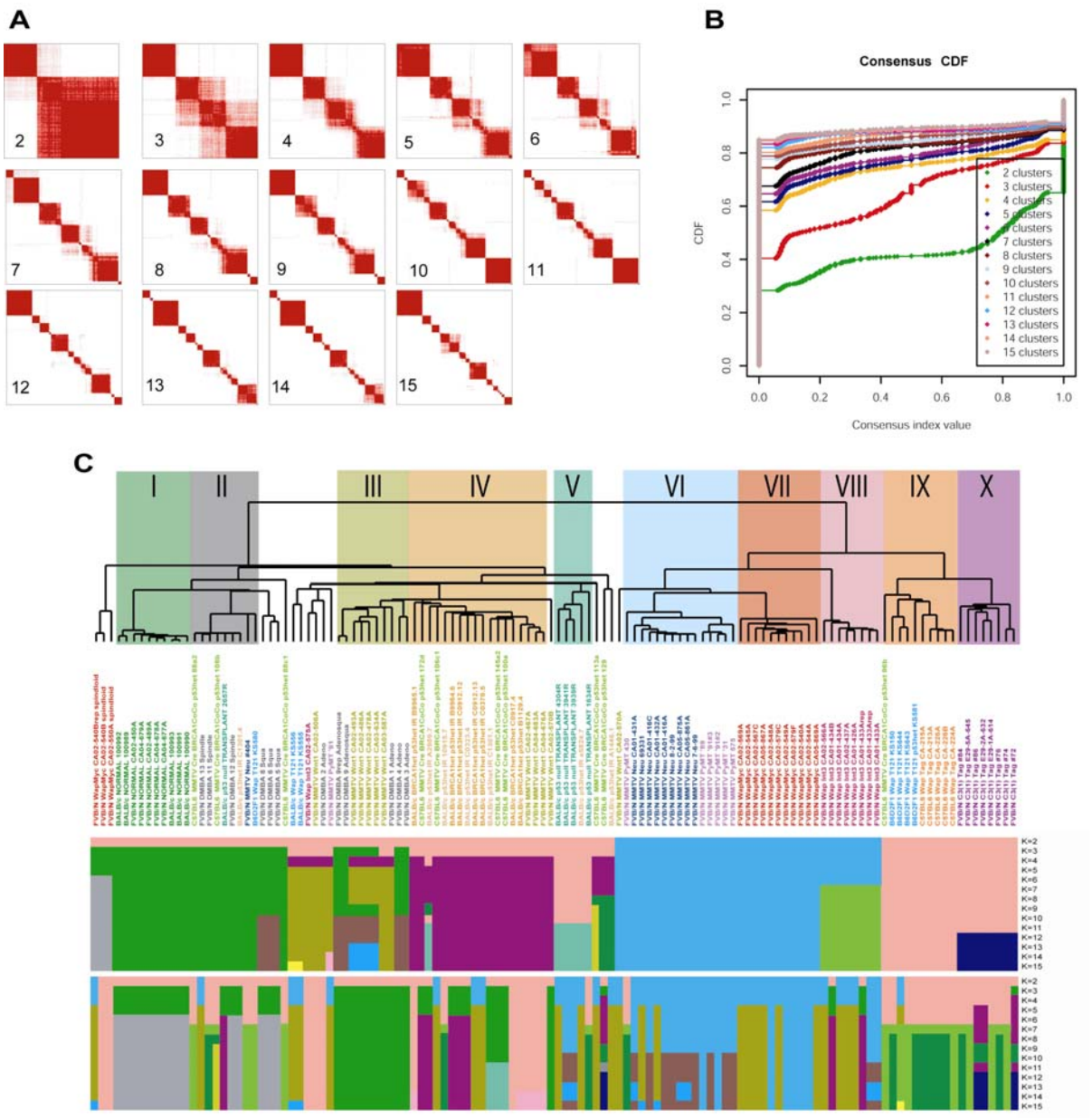
39. Mosley JD, Poirier JT, Seachrist DD, Landis MD, Keri RA: **Rapamycin inhibits multiple stages of c-Neu/ErbB2 induced tumor progression in a transgenic mouse model of HER2-positive breast cancer**. *Mol Cancer Ther* 2007, **6**(8):2188-2197.

APPENDIX IIA



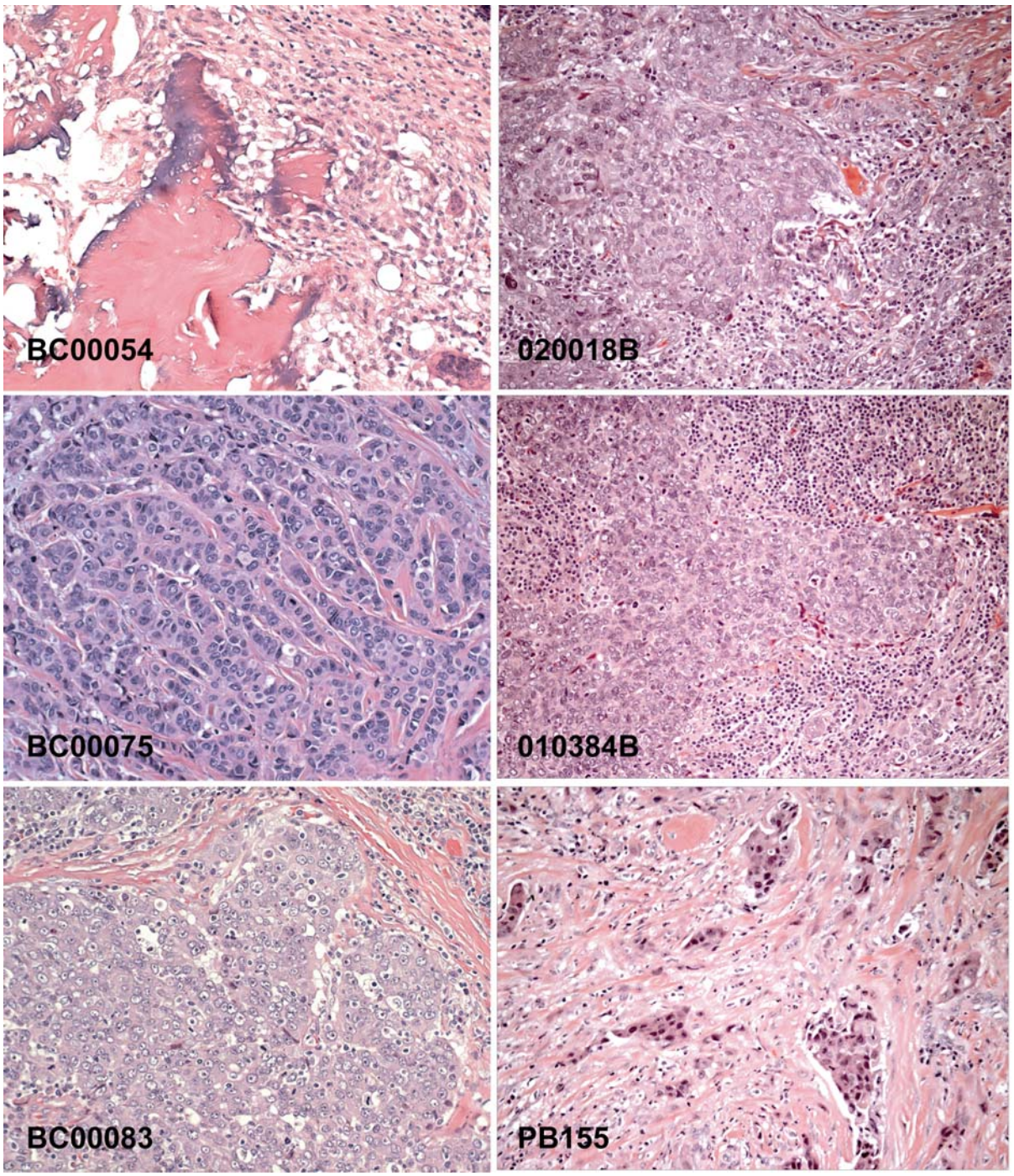
Appendix IIA. Complete unsupervised cluster diagram of all mouse tumors. Samples are colored according to mouse model from which they were derived, and the genes were selected using a variation filter of 3-fold or more on 3 or more samples.

APPENDIX IIB



Appendix IIB. Consensus Clustering (CC) analyses applied to the mouse models. (A) CC matrices generated using the 866 gene mouse intrinsic list, by cluster numbers K=2 through K=15. (B) Empirical Cumulative Distribution (CDF) plot corresponding to the consensus matrices in the range K=2 to 15. (C) CC directly compared to the hierarchical clustering-based results. The dendrogram from Figure 1 (using the intrinsic gene set) is shown and immediately below is a colored matrix showing sample assignments based upon the various number of K clusters from the CC. By comparison, the analysis performed on the mouse dataset using all genes (bottom matrix) is presented.

APPENDIX IIC



Appendix IIC. Histological characterization of six different human "claudin-low" tumors using H&E sections.

APPENDIX IID

Appendix IID. Gene Set Enrichment Analysis (GSEA) of murine pathway models versus 5 human subtypes. Statistically significant findings are highlighted in bold.

Is Class

		BASAL-like		LUMINAL		HER2+/ER-		NORMAL		claudin-LOW	
Mouse Model	# genes	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val
T-antigen models	406	0	**0.002**	-	-	-	-	-	-	-	-
BRCA1 models	427	0.004	0.0881	-	-	-	-	-	-	-	-
Wnt1 model	34	0.1159	0.5215	-	-	0.1741	0.6356	0.0629	0.2553	-	-
Myc.model	517	0.1438	0.6086	-	-	0.4553	0.948	-	-	-	-
BRCA1 p53 IR model	65	0.1957	0.6867	-	-	-	-	0.3354	0.8739	-	-
Int3 model	137	0.283	0.8428	-	-	0.1706	0.7417	0.0278	0.3904	-	-
Neu.model	460	-	-	0.0061	0.154	0.1594	0.6847	0.9014	0.997	-	-

Is Not Class

		BASAL-like		LUMINAL		HER2+/ER-		NORMAL		claudin-LOW	
Mouse Model	# genes	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val
T-antigen models	406	-	-	0.043	0.087	0.899	0.999	0.06831	0.207	0.5962	0.996
BRCA1 models	427	-	-	0.1037	0.52	0.5203	0.968	0.14	0.605	0.7519	0.997
Wnt1 model	34	-	-	0.01734	0.113	-	-	-	-	0.2417	0.787
Myc.model	517	-	-	0.126	0.557	-	-	0.7692	0.998	0.278	0.85
BRCA1 p53 IR model	65	-	-	0.5198	0.966	0.6536	0.991	-	-	0.1063	0.499
Int3 model	137	-	-	0.1996	0.799	-	-	-	-	0.2079	0.83
Neu.model	460	0	**0.027**	-	-	-	-	-	-	0.4188	0.939

APPENDIX IIE

Appendix IIE. Gene Set Enrichment Analysis (GSEA) of 10 murine classes versus clinical ER status and HER2 status in ER negative patients. Statistically significant findings are highlighted in bold.

Is Class

		ER +		**ER- / HER2-**		**ER- / HER2+**	
Mouse Class	# genes	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val
I	1882	0.2267	0.795	-	-	0.9243	0.997
II	912	-	-	-	-	0.5675	0.988
III	143	0.7919	0.999	-	-	0.2723	0.837
IV	1019	-	-	0.0062	0.323	-	-
V	34	-	-	0.668	0.994	0.4929	0.964
VI	820	0.0098	0.174	-	-	0.6111	0.976
VII	851	-	-	0.1417	0.618	0.5666	0.975
VIII	236	0.3379	0.922	-	-	0.1244	0.73
IX	462	-	-	0	**0.009**	0.8112	0.998
X	338	-	-	0	**0.003**	-	-

Is Not Class

		ER +		**ER- / HER2-**		**ER- / HER2+**	
Mouse Class	# genes	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val
I	1882	-	-	0.0829	0.522	-	-
II	912	0.7182	0.999	0.7589	1	-	-
III	143	-	-	0.5337	0.978	-	-
IV	1019	0.012	0.469	-	-	0.3107	0.909
V	34	0.5571	0.985	-	-	-	-
VI	820	-	-	0.0041	0.061	-	-
VII	851	0.1726	0.675	-	-	-	-
VIII	236	-	-	0.2432	0.873	-	-
IX	462	0.00641	**0.024**	-	-	-	-
X	338	0.00641	**0.043**	-	-	0.6922	0.999

APPENDIX IIF

Appendix IIF. Gene Set Enrichment Analysis (GSEA) of murine pathway models versus clinical ER status and HER2 status in ER negative patients. Statistically significant findings are highlighted in bold.

Is Class

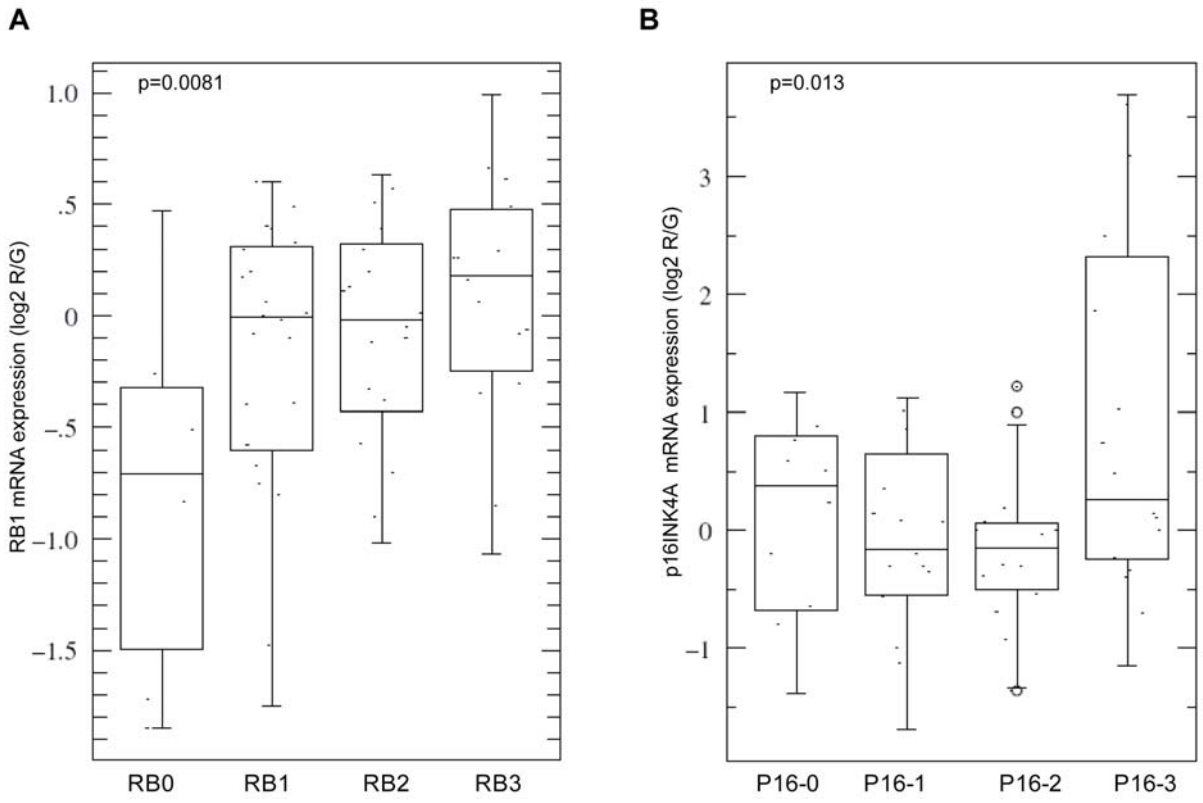
		ER +		**ER- / HER2-**		**ER- / HER2+**	
Mouse Model	# genes	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val
T-antigen models	406	-	-	0	**0.003**	-	-
BRCA1 models	427	-	-	0.0061	0.116	0.7755	0.993
Wnt1 model	34	-	-	0.4073	0.938	0.2478	0.775
Myc.model	517	-	-	0.1253	0.573	0.5506	0.975
BRCA1 p53 IR model	65	-	-	0.6159	0.993	0.5648	0.971
Int3 model	137	0.1988	0.818	-	-	-	-
Neu.model	460	0.0078	0.185	-	-	0.5479	0.965

Is Not Class

		ER +		**ER- / HER2-**		**ER- / HER2+**	
Mouse Model	# genes	NOM p-val	FWER p-val	NOM p-val	FWER p-val	NOM p-val	FWER p-val
T-antigen models	406	0.0106	**0.022**	-	-	0.8453	1
BRCA1 models	427	0.0098	0.17	-	-	-	-
Wnt1 model	34	0.3252	0.858	-	-	-	-
Myc.model	517	0.1206	0.604	-	-	-	-
BRCA1 p53 IR model	65	0.6063	0.985	-	-	-	-
Int3 model	137	-	-	0.3202	0.899	0.5552	0.978
Neu.model	460	-	-	0.002	0.079	-	-

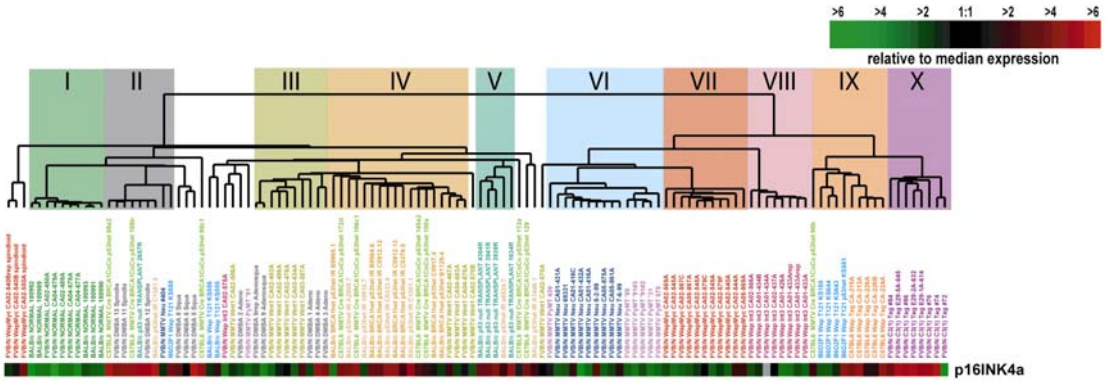
APPENDIX VA

Appendix VA. ANOVA box plot comparison of mRNA expression and protein staining in breast tumors. A) shows low RB1 mRNA expression correlates with low RB1 protein expression and B) high p16INK4a mRNA expression correlates with high protein expression.



APPENDIX VB

Appendix VB. p16INK4a expression in mouse models of breast cancer.



APPENDIX VC

Appendix VC. Comparison of overlap between 3 gene lists using hypergeometric mean analysis.

Overlap of population of RB-LOH and RB-regulated-inclusive: 7722 genes
RB-LOH in the overlap: 270 genes
RB-regulated-inclusive in the overlap: 119 genes
Overlap of RB-LOH and RB-regulated-inclusive in the overlap of the population: 42 genes

phyper(42,270,(7722-270),119)
p-value < 0.001

++

Overlap of population of RB-regulated-inclusive and V12: 8381 genes
RB-regulated-inclusive in the overlap: 115 genes
V12-Prolif in the overlap: 119 genes
Overlap of RB-regulated-inclusive and V12-Prolif in the overlap of the population: 27 genes

phyper(27,115,(8381-115),119)
p-value < 0.001

++

Overlap of population of RB-LOH and V12: 10407 genes
RB-LOH in the overlap: 329 genes
V12-Prolif in the overlap: 137 genes
Overlap of RB-LOH and V12-Prolif in the overlap of the population: 68 genes

phyper(68,329,(10407-329),137)
p-value < 0.001

APPENDIX VD

Appendix VD. The expression of RB-loss signature is highest in basal-like tumors.